Educational and Developmental Outcomes in England: The Impact of Formal Childcare Intensity, School Absences and Peer Ethnicity

Lucy Ward

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Sheffield
School of Economics

September 2024

Abstract

This thesis consists of three separate empirical studies. Using individual-level data from administrative and survey sources, as well as adopting econometric techniques, this thesis examines the determinants of individuals' cognitive and non-cognitive skills and beliefs about future education outcomes in England.

Chapter 2 uses data from the Millennium Cohort Study and applies an instrumental variables approach, exploiting exogenous variation in the likelihood that the mother works shift work or has uncertain working hours. This strategy estimates the impact of formal childcare hours before age 3 on non-cognitive skills between ages 3 and 14. The findings suggest that increased time in formal childcare has a positive initial effect on non-cognitive skills, with these benefits persisting over time.

Chapter 3, using administrative pupil-level data matched with longitudinal survey responses, quantifies the impact of school absences on Maths and English end-of-year exam results in a UK context. Exploiting the panel structure of the data, by using multi-dimensional fixed effect analysis, we find that absences significantly reduce academic performance.

Chapter 4 examines the impact of ethnic minority peers on white pupils' perceived likelihood at age 14 of attending university. Matching the Millennium Cohort Study to the National Pupil Database (NPD) and creating an instrumental variable strategy based on nurse recruitment in 1949, our results indicate that increasing the proportion of ethnic minorities in the school positively and significantly impacts white pupils' perceived likelihood at age 14 of attending university.

This thesis provides policy-relevant insights to inform effective educational strategies. The studies underscore the importance of expanding access to formal childcare, addressing school absences, and enhancing ethnic diversity within schools. The findings are relevant to other

social scientists, teachers, and parents.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisors, Steve McIntosh and Bert Van Landeghem, for their unwavering support and guidance throughout this process. The completion of this thesis would not have been possible without your help in navigating the various hurdles of this project. Thank you for believing in me.

I am grateful to the School of Economics for the scholarship that I received to conduct my PhD studies. Additionally, I would like to express my gratitude to my fellow PhD students, both past and present—Tom, Michael, Maria, and Adriana. It has been a great experience to share this journey with you. I am thankful to all the academic and support staff in the School of Economics for their help and support throughout this process. I would like to specifically mention Tom Siddall and Andy Bryce for providing support and friendship, which has been a source of reassurance.

I would also like to say a heartfelt thank you to my family—my dad, Damian, my mum, Melanie, and my siblings, Emily and Oscar—for their endless support, love, and encouragement. Finally, but most importantly, to my boys, Jake and Fudge, these last three years would not have been possible without you.

Notes and Disclaimers

I confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not previously been presented for an award at this, or any other, university.

Chapter 3 and Chapter 4 of this thesis are based on data collected from the National Pupil Database (NPD) which is controlled by the Department for Education (DfE) and a secure version of the Millennium Cohort Study (MCS). This data is supplied by the Secure Data Service (SDS) at the UK Data Archive.

The use of this data does not imply the endorsement of the data owner or the UK Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

An edited version of Chapter 2 has been published in Sheffield Economic Research Paper Series (SERPS).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Improving the educational outcomes of children and young people is an objective held by political parties across the world. Education can drive development at an individual and country level. Education can lead to improved life conditions and a reduction of inequalities. In addition, research has shown education to be related to increases in the probability of good health (Feinstein 2002), happiness (Blanchflower & Oswald 2004), reduced population growth (Becker et al. 1990), crime (Lochner 2020), and poverty (Biosca et al. 2014). Understanding the determinants of educational outcomes is crucial not only for enhancing individual life prospects but also for fostering societal well-being and growth.

Education expenditure accounts for relatively high shares of governments' total expenditure. Across the EU, government expenditure on education averaged 9.5% in 2022 (European Commission 2024). The UK government spent £116 billion on education in 2022 (Drayton et al. 2023). High spending on education drives the need for evidence informed policy decisions. However, due to significant gaps in the literature, challenges remain in making this a reality.

This thesis focuses on England, which is recognised for its high school-age attainment. Over the last 15 years, literacy and numeracy skills in England have improved significantly in comparison to other high-income countries. Among the OECD only Canada, Estonia, Ireland and Japan manage to deliver both stronger average attainment and lower inequality than England (Farquharson et al. 2024).

Despite this, there are still significant inequalities within education outcomes in England. Farquharson et al. (2022) show that at age 5, only half of pupils eligible for free school meals (FSM) achieve the recommended level of development, compared with 72% of non-FSM-eligible pupils.[1] Further on in the education journey, 28% of pupils eligible for FSM attend university compared with 47% of non-FSM-eligible pupils. Conditional on attending university, pupils eligible for FSM are three times less likely to attend a more selective university. Despite large policy attention, many education inequalities have persisted over time. For example, there has been very little change in the gap between children from more and less disadvantaged backgrounds in GCSE attainment over the past 20 years. GCSE attainment has been increasing over time, however, only 41% of pupils eligible for FSM pass English and maths compared to 69% of all other pupils.

This thesis addresses three key policy areas: childcare, absences, and ethnic diversity. Addressing these issues could provide pathways to mitigate educational inequalities and enhance overall outcomes. Below we set out the motivation behind each of these areas of focus.

Firstly, whilst 72% of non-disadvantaged pupils reach the expected level of development by the age of 5, only 55% of disadvantaged pupils reach the same level in the UK (Andrews

[1]The use of free school meal eligibility as a proxy for being disadvantaged has two main limitations. Firstly, it may not identify some groups, such as children in 'working poor' households (Hobbs & Vignoles 2010). Additionally, free school meal eligibility is linked to parental employment which can change through a pupils time at school. Other measures of disadvantage, for example low parental education, are persistent (Ilie et al. 2017).

et al. 2017). Disparities in skills during early childhood are likely to contribute to inequalities later in life, including health (Case et al. 2005), education (Almond & Currie 2011), and employment (Black et al. 2007). The UK Conservative government's commitment to halve this gap by 2028, has created a conversation on how to achieve this goal. Early childhood has become a central focus in this discussion, particularly in relation to the development of human capital.

Despite the recognised importance of early childhood in the human capital production function, it remains the only stage in childhood and adolescence with relatively limited public investment. In the UK, less than 0.1% of GDP is allocated to childcare, placing it among the lowest investments within the OECD (OECD 2023).

In the 2023 budget, the Conservative government announced their plans to provide 30 hours of free childcare for children over the age of nine months with working parents. The Labour government have said they would not reduce this new entitlement to free childcare.

Secondly, absence from school is now one of the most pressing issues facing England's education system. Gillian Keegan, the last Conservative Education Secretary said attendance was her number one priority. This focus is unlikely to change under the Labour Government as in their election campaign they described absences as a generational challenge.

Whilst it was assumed that school absences would quickly fall back to their pre-Covid level, absence rates continue to be high four years later. In England, the absence rate for 2018/19 was 4.7%, which equates to 59.6 million days lost (Department for Education 2019). In 2022/23, absences rose by 60% to 7.5% of school days being missed (Eyles et al. 2023). We are now not only seeing persistently high levels of absences, but they are also not as socially stratified as prior to the pandemic. Most pupils are absent more often, post-pandemic than pre. We are now potentially facing a worldwide education crisis in the post-pandemic era: the majority of pupils are not maximising the quantity of instructional time. With a rise in absences across the pupil population, it is now more important than ever to quantify the causal impact of absences on achievement.

Finally, many places in Britain are now being described as ethnically diverse (Catney et al. 2023). Whilst ethnic minority pupils start school behind white pupils, they make faster progress through the education system. All major ethnic groups are more likely than white pupils to achieve A levels or equivalent qualifications and, on average, significantly more likely to go to university (Farquharson et al. 2022). In the 2022 Programme for International Student Assessment (PISA) the UK was the only country in Europe where second-generation immigrants outperform non-immigrant students in maths (Ingram et al. 2023).

The impacts of immigration on various outcomes remains a highly debated and controversial topic across the world. Census data for England and Wales shows that, from 2001 to 2021, the percentage of people who identified as White British reduced from 87.5% to 74.4%, highlighting the importance of understanding the impacts of ethnic diversity (Office of National Statistics 2024).

Through causal research in these areas, this thesis aims to contribute to a more fair and effective educational landscape in England.

In the following section, the aims and contributions of this thesis are highlighted, the research questions addressed are set out and an overview of each chapter is presented.

1.2 Aims, research questions and contribution of this thesis

This thesis consists of three related yet standalone empirical studies exploring the determinants of individuals' cognitive and non-cognitive outcomes and beliefs about future education outcomes in England.

Each chapter examines administrative and survey data to provide robust empirical evidence to address the research questions. To build on the empirical literature in this field, the following research questions are addressed by this thesis:

- Chapter 2:
  1. What is the causal effect of hours spent in formal childcare at age 3 on non-cognitive skills from age 3 to 14?
  2. Does the impact of hours spent in formal childcare differ by type of non-cognitive skill?
  3. Is time spent in childcare more beneficial for disadvantaged children?
- Chapter 3:
  1. What is the causal effect of school absences on end of year achievement in maths and English?
  2. Is the relationship between absences and academic achievement non-linear?
  3. Does the impact of absences differ depending on if they are authorised or unauthorised?
  4. Does the impact of absences differ depending on the timing?
- Chapter 4:
  1. What is the causal effect of peers' ethnicity on the perceived likelihood of attending university for white pupils aged 14?
  2. Do peer effects on university likelihood vary by gender and family background?
  3. Do the peer effects differ depending on which ethnic groups a pupil is exposed to?

This thesis contributes to the Economics of Education literature in three significant ways. Firstly, we provide novel evidence on the impact of time in formal childcare, absences, and peer ethnicity in influencing individual outcomes. All these areas are of current policy relevance and are impacted by the lack of available evidence. Secondly, we use both administrative and survey data allowing us to control for many confounding factors, particularly in Chapter 3 where we can control for time-varying characteristics. Finally, we make methodological contributions by creating new instrumental variables in Chapters 2 and 4.

All three empirical studies presented in this thesis focus on the English schooling system. The following section describes the compulsory English education system to give institutional context.

1.3 Institutional background: The English Education System

Education in England is overseen by the UK's Department for Education. The 152 Local Education Authorities (LEAs) in England are responsible for the provision of education at the local level. They set rules on admission, teacher pay, curriculum, school term dates, etc. England also has a number of academies that run independently from the LEAs.

65.5% of schools in England are run by local authorities (Department for Education 2024). These schools take both boys and girls in mixed classes. They also admit pupils without reference to ability and cater for all the children in a neighbourhood. In some areas, they co-exist with other types of public schools, for example, grammar schools which admit pupils based on ability. There are also academies which are run independently from local authorities, private schools which are fee paying schools and some parents choose to home school their children.[2] Ofsted is a non-ministerial department which is responsible for inspecting schools in England.

Education became compulsory in 1870. Full-time education is now compulsory from age 5. Children start school the September after their 4th birthday.[3] In 1972, the education participation age was raised from 15 to 16. In 2008 it was announced that the school leaving age was once again going to be increased. In 2013, the education participation age was increased to 17, followed by 18 in 2015.

Compulsory education in England is made up of three main parts, primary education, secondary education, and further education. The school curriculum is organised in blocks of years called Key Stages (KS). Pupils enter school in Reception and remain there for a year before moving to KS1 until aged 7, they then move to KS2 between the ages of 7-11.[4] When pupils are 11, they leave the primary stage and move to secondary education. This normally involves moving schools. During the primary stage, pupils focus on Maths, English, and Science. When they move to secondary school, they study a wider range of subjects including History, Geography, Languages, and Religious Education to name a few. Pupils begin secondary school in KS3 and remain there until age 14, when they move to KS4 to complete their GCSEs or equivalent exams at age 16.

Prior to 2013, pupils could leave the education system after completing their GCSEs, though many choose not to. After the reforms to the school leaving age in 2013, pupils had to stay in education or recognised training. This means pupils move to Key Stage 5, to complete their A-levels or equivalent.[5] Pupils either stay at a school that has a sixth form or go to college. Pupils also have the option of learning as an apprentice.

[2] Green (2024) suggests that 6.4% of England pupils are privately educated.

[3] Some pupils will be nearly 5 by the time they start school whilst others will have just turned 4.

[4] KS1 and KS2 are normally completed in the same school.

[5] Some pupils will still be taking GCSE courses for example, if pupils fail Maths and/or English GCSE then they must retake these courses when they enter Key Stage 5.

A breakdown of the English schooling system is provided in Table 1.1.

Table 1.1: The English Schooling System

Type of school	Key Stage	Year group	Age	Exams
Primary	Early Years	Reception	Age 4-5	
	1	Year 1	Age 5-6	
		Year 2	Age 6-7	Key Stage 1 SATs
	2	Year 3	Age 7-8	
		Year 4	Age 8-9	
		Year 5	Age 9-10	
		Year 6	Age 10-11	Key Stage 2 SATs
Secondary	3	Year 7	Age 11-12	
		Year 8	Age 12-13	
		Year 9	Age 13-14	Key Stage 3 SATs
	4	Year 10	Age 14-15	
		Year 11	Age 15-16	GCSEs and equivalents
Further education	5	Year 12	Age 16-17	
		Year 13	Age 17-18	A-level and equivalents

1.4 Structure and content of this thesis

This thesis is made up of three separate empirical studies presented in Chapters 2, 3 and 4. Each of these studies uses individual-level data from administrative records and surveys, employing econometric techniques to examine the determinants of individuals' cognitive and non-cognitive outcomes and beliefs about future education outcomes in England. Chapter 5 concludes this thesis. The three chapters forming the main part of this thesis are briefly discussed below.

1.4.1 Brief overview of Chapter 2

Chapter 2 focuses on inventions prior to school and examines the impact of time spent in formal childcare by age 3 on non-cognitive skills both in the short and long run.

We use data from the Millennium Cohort Study, and an instrumental variables strategy that leverages exogenous variation in both the probability that the mother works shift work and has uncertain working hours based on the job stated during pregnancy, as instruments for childcare usage.

We find that increasing hours in formal childcare has an initial positive impact on non-cognitive skills which persists over time. The impact on emotional skills, conduct and peer relationships could be argued to be the most persistent over time. Moreover, we estimate heterogeneous impacts across family background characteristics, suggesting that increasing access to more time in childcare for disadvantaged children may hold potential for decreasing early inequalities in child development. The results are robust to a number of sensitivity checks including weak instrument robust testing and discussion of potential omitted variables.

1.4.2 Brief overview of Chapter 3

Chapter 3 uses the National Pupil Database (NPD) which is an administrative pupil level dataset matched with Understanding Society (UKHLS)- Harmonised British Household Panel data (BHPS) a longitudinal survey, to examine the relationship between school absences and end of year achievement in maths and English in England. Raising school attendance has long been a key policy focus, and the Covid-19 pandemic has further intensified the issue of school absenteeism.

Exploiting the panel structure of the data, by using multi-dimensional fixed effect analysis, whilst additionally controlling for potential time varying factors such as household income, we find that absences significantly reduce academic performance. We find approximately linear effects of pupil absences on end of year achievement in Maths and English with no evidence of a discontinuity at the persistently absent threshold. We also provide evidence to suggest that reducing the absences of those pupils at the bottom of the ability distribution could aid in reducing the performance gaps.

Our findings deliver very robust results across different samples of individuals and multiple measures of achievement. The robustness checks consistently suggest that the existence of omitted variable bias, reverse causality or measurement error are unlikely, which supports a causal interpretation. The findings are also comparable to the literature.

1.4.3 Brief overview of Chapter 4

Chapter 4 uses the Millennium Cohort Study, a nationally representative English birth cohort which is linked with the NPD to investigate whether the reported likelihood at age 14 of attending university for white pupils is influenced by the ethnicity of their peers.

To address the endogeneity and selection biases encountered when estimating peer effects, we developed a novel instrumental variable. We identify exogenous variation in the ethnic composition of a school by exploiting the government-backed recruitment drive for nurses from the Commonwealth, that began in 1949. Using occupation records from 1951 to identify areas of England that experienced nurse shortages, we use the proportion of nurses in the local area as an instrument for the proportion of ethnic minority pupils within a school in 2014.

Our results suggest that increasing the proportion of ethnic minorities in the school has a positive and significant impact on white pupils' perceived likelihood at age 14 of attending university. Moreover, we estimate homogenous impacts across gender with some evidence of heterogeneous impacts across family background characteristics, suggesting that improving ethnic diversity within schools could aid in improving higher education participation for underrepresented groups. The results are robust to a number of sensitivity checks including weak instrument robust testing, changes in the sample and potential omitted variables.

Chapter 2

Unlocking Potential: Investigating the Prolonged Impact of Formal Childcare Intensity on Non-Cognitive Skills

2.1 Introduction

Are we falling short in providing adequate support for our children before they start school? Whilst 72% of non-disadvantaged pupils reach the expected level of development by the age of 5, only 55% of disadvantaged pupils reach the same level in the UK.[1] (Andrews et al. 2017). Disparities in skills during early childhood are likely to contribute to inequalities later in life, including health (Case et al. 2005), education (Almond & Currie 2011), and employment (Black et al. 2007). The UK government's commitment to halve this gap by 2028, has created a conversation on how to achieve this goal. Early childhood has become a central focus in this discussion, particularly in relation to the development of human capital.

[1] Disadvantaged is defined as being eligible for free school meals.

There is a wide range of literature on the benefits of attending childcare, with specific focuses on the benefits for disadvantaged children. Less research has focused on the intensity of this treatment. This chapter aims to shed light on the causal relationship between hours spent in childcare by age 3 and non-cognitive development for children aged 3-14. Using nationally representative data from the Millennium Cohort Study, we are able to show that there are not only short run gains but also long run impacts of early access to childcare for more hours.

Despite the recognised importance of early childhood in the human capital production function, it remains the only stage in childhood and adolescence with relatively limited public investment. In the UK, less than 0.1% of GDP is allocated to childcare, placing it among the lowest investments within the OECD. Notably, 84% of childcare in the UK is provided by for-profit providers, a sharp contrast to the 3% in Germany and 4% in France (Menon 2023). The UK childcare system is also much more heterogeneous than in other European countries (Sweden, Denmark, and France), where an affordable, full time public childcare system has been in place for several decades. The cost of childcare in the UK has been notoriously high with the government first introducing childcare subsidies in the early 2000s.[2]

Allocating resources to childcare could create numerous advantages, including the enhancement of children's development, the increase in family income, and a general improvement in productivity. This is because childcare is likely having knock-on implications for parents' ability and incentive to work. In the short term, it boosts productivity by expanding labour force participation and hours of work. Additionally, there are medium-term benefits, as childcare prevents the deterioration of parents' skills during periods of non-employment. Over the long term, investing in childcare is expected to contribute to sustained productivity

[2]All 4-year-olds since 2000 and all 3-year-olds since 2004 have been entitled to free, part-time nursery education. Starting in 2000, the Department for Education funded childcare places for three-year-olds in 65 Local Education Authorities (LEAs). By 2001, this initiative was expanded nationwide, with the goal of achieving universal coverage for all three-year-olds by 2004. This was set at 12.5 hours per week for 33 weeks per year until 2010, then 15 hours per week for 38 weeks. Sure Start began in 1999 but did not offer childcare until 2004. In 2013, free part-time education was also extended to disadvantaged 2 year olds. In 2017, this was increased to 30 hours for working parents. In the 2023 budget, the government announced their plans to provide 30 hours of free childcare for children over the age of nine months with working parents.

by improving children's development.

Previous literature has provided international evidence on the positive impacts of attending childcare prior to starting school. This research has demonstrated both short run and long run impacts on cognitive and non-cognitive development (see Elek et al. (2020) for an overview of the literature). Areas that have received less attention in the literature are on the effectiveness of attending childcare for children under 3 and the intensity of childcare. What evidence there is, suggests that starting childcare at a younger age and for more hours provides positive impacts (Del Boca et al. 2018, Felfe & Lalive 2018, Drange & Havnes 2019, Morris et al. 2021A). Note, however, these findings are not conclusive and are dependent on individual and family characteristics as well as care settings (Gupta & Simonsen 2010, Kuehnle & Oberfichtner 2020).

This research aims to explore the relationship between hours in childcare prior to the age of 3 and non-cognitive skills, an area which has received limited attention in the literature. In this chapter, we use the term non-cognitive skills to describe the personal attributes not thought to be measured by achievement tests.[3] Non-cognitive skills are often overlooked within the education assessment system where there is heavy reliance on written tests to screen and sort individuals, to evaluate pupils and schools, and to assess the performance of entire nations. This is despite the argument that non-cognitive skills predict a wide range of outcomes, including educational achievement, labour market outcomes, health, and criminality (Heckman & Kautz 2014, Almlund et al. 2011, Borghans et al. 2008, Roberts et al. 2007). Furthermore, the predictive power of non-cognitive skills has been shown to rival that of measures of cognitive ability. For example, conscientiousness predicts years of schooling with the same strength as measures of intelligence (Almlund et al. 2011).

Against this backdrop, we make three contributions to the literature. Firstly, we add to

[3]These attributes go by many names in the literature, including soft skills, personality traits, non-cognitive abilities, character skills, and socio-emotional skills.

the knowledge base on the impacts of childcare on non-cognitive skills. Due to variation in findings, providing useful evidence for policy guidance is a challenge, and by increasing the evidence there will be a stronger evidence base to underpin conclusions. Secondly, this research specifically focuses on the impact of hours in childcare. This is the first research to shed light on the causal relationship between hours in childcare and non-cognitive skills both in short and long run. In the UK, for example, the current childcare policy focus is on the number of hours which should be subsidised though the evidence base on this question is limited. Finally, we focus on childcare taken at ages up to age 3, while the majority of the literature has examined the impact of childcare on children aged 3-5. We think the focus on this younger age group is important for two main reasons. Firstly, in the 2023 budget, the government announced their plans to provide 30 hours of free childcare for children over the age of nine months with working parents. The policy focus is therefore currently on this younger age group for which the literature is unclear on the impact of childcare. Secondly, the Heckman Curve demonstrates that the highest economic returns are achieved through early investments in children (Heckman 2011).

The economics literature supports the idea that investments in children's skill development should be focused at earlier stages, rather than later in their lives (Heckman & Kautz 2014). Investments made during the preschool years are less costly and more effective. This is because young children's abilities are more adaptable and the benefits accumulate over time. Spending time away from their parents as well as being exposed to various peers at a young age may foster independence and development of socio-emotional skills. The amount of time spent in childcare may also matter. Despite research demonstrating the positive impacts of childcare, the 'attachment theory' (Bowlby et al. 1951) stresses the importance of the bond formed between children and their parents at young ages. This may suggest that there is an optimal amount of time spent in childcare which creates a balance between time spent with and away from their parents.

As this research focuses on children under 3 years olds in 2003/4, who were not eligible for the childcare policies of the time, this implies that any formal childcare observed is mainly determined by parents' choices. Identifying a causal relationship is therefore a challenge due to unobserved confounding factors. Parents' parenting skills and preferences for childcare are likely correlated with the amount of time spent in childcare and the child's non-cognitive skills. These factors are difficult to observe, potentially creating a spurious correlation between hours in childcare and non-cognitive development. To correct for endogeneity of hours in childcare, this research creates and uses instrumental variables.

Making use of the information on mother's occupation during pregnancy, we calculate the probability that the mother would work shifts or have uncertain working hours. The assumptions are that due to the structured nature of formal childcare, mothers working in sectors with more shift work, place their child in formal care for fewer hours. On the other hand, if the mother has varying hours each week, the uncertainty causes them to place their child in formal care for more hours.

Using the Millennium Cohort Study (MCS), which provides detailed information about childcare and child outcomes, our results are summarised as follows: the amount of time spent in formal childcare has a positive and significant impact on non-cognitive skills at age 3 through to age 7. We find that these effects are larger for disadvantaged children. The findings are robust to a number of specification tests including weak instrument robust testing, changes in the sample and potential omitted variables. This research adds to the evidence-based case for significant additional government investment in early childcare.

This research is organised as follows. Section 2.2 reviews the existing literature. Sections 2.3 and 2.4 outline the data and methodology respectively. Section 2.5 presents the estimated impact and tests the robustness, before concluding in Section 2.6.

2.2 Literature Review

Research on the effects of childcare on child outcomes has predominantly focused on evaluating small-scale interventions aimed at low-income families in the US. These programs (High Scope/Perry Preschool, Infant Health and Development Program) have been described as providing high-quality centre-based care where the quality is enforced and monitored. Additional to providing childcare, they also offer parent focused support such as information sessions and home visits. Research has shown large long-term positive effects that span both cognitive and non-cognitive outcomes (Belfield et al. 2006). Due to the specific nature of these policies, generalisability is difficult. Less research has been undertaken, on what Melhuish (2004) describes as 'general population research', to understand the impact of types of childcare experienced by the majority of children.

Only a few studies have used large datasets specifically gathered on preschool-aged children, such as the Effective Provision of Pre-School Education (EPPE) in the UK and the National Institute of Child Health and Human Development (NICHD) in the US. In both cases, preschool children were recruited and tracked over several years. Research carried out using both datasets finds positive effects of childcare on test scores in both the short and long run (Waldfogel 1999, Sylva et al. 2004, Vandell et al. 2010). The estimated impact on non-cognitive outcomes is slightly more mixed. Waldfogel (1999), using US data, finds that children who spent more time in childcare were found to have more behavioural problems, whilst Sylva et al. (2004), using UK data, finds a positive impact of childcare on social development. Whilst these studies provide a better understanding of the wider impacts of childcare, with large changes in childcare policy in recent years, these data could be out of date as NICHD was created in 1991 whilst EPPE was launched in 1997.

More recent studies have used quasi-experimental methods to evaluate universal provision of childcare. Many studies have leveraged variation in the timing of policies (Felfe & Lalive

2018, Felfe et al. 2015), while others have created instrumental variables (Berger et al. 2021, Gupta & Simonsen 2016). Such studies have produced diverging estimates, from positive (Berger et al. 2021, Kuehnle & Oberfichtner 2020, Fort et al. 2020, Carta & Rizzica 2018, Del Boca et al. 2016, 2018, Felfe & Lalive 2018, Caprera 2015, Blanden et al. 2016, Gupta & Simonsen 2010) to negative (Baker et al. 2019, Magnuson & Duncan 2016, Burger 2010) and insignificant (Kuehnle & Oberfichtner 2020, Hansen & Hawkes 2009).

Within this literature review we will discuss some of the most recent studies on the impact of childcare on child outcomes. We begin by presenting some of the theoretical explanations for the impact before highlighting the identification issues empirical analysis faces. For the main part of the review, we critically evaluate the methods used by previous studies to try to overcome some of these issues. We end the review by stating the contribution of this research.

2.2.1 Theory

Theoretical explanations on the impact of childcare are mixed.

A significant portion of the economic literature on the relationship between childcare and child outcomes has been shaped by the work of Todd & Wolpin (2003) and James Heckman and his co-authors (e.g., Carneiro & Heckman (2003)). They modeled children's outcomes as the result of a production function, where inputs are provided not only by families but also by other individuals and institutions, such as schools, teachers, peers, and society. Within this framework, child development is viewed as the outcome of a cumulative process of knowledge acquisition. They argue that the highest economic returns come from the earliest investments in children, as these early investments are both less costly and more effective. This effectiveness is due to the greater malleability of young children's abilities, which have a cumulative impact over time, unlike investments made later in life.

An opposing theory comes from the psychological literature where the attachment theory (Bowlby et al. 1951) stresses the importance of the bond formed between children and their caregivers at young ages. Bowlby et al. (1951) argues that the earliest bond formed by children with their caregivers has a large impact that continues throughout their life. Theoretically, non-maternal childcare may be negatively associated with child development because it reduces the time a child spends with their mother, potentially hindering bonding and possibly slowing brain development.

2.2.2 Identification issues

Unlike primary education, in the UK, pre-school care is non-compulsory meaning parents choose whether to send their child. The impact of childcare will therefore only be correctly estimated through OLS if the selection on observables assumption is satisfied. This means that all variables which predict both childcare and children's outcomes must be included in the model.

As it is normally the parents' decision to place their child in childcare, we must consider the ways parents affect their children. Parents with strong parenting skills provide warm, supportive, and nurturing environments for their children, regardless of their financial resources, the amount of time they spend together, or specific child development outcomes. Those with greater financial and time resources have the ability to invest more in goods (e.g., tuition) and activities (e.g., library visits), though whether they choose to do so depends partly on their preferences. These preferences are only implied through actions and likely unobserved. The decision to use childcare may also depend on children's characteristics such as behaviour or natural ability. The direction of the bias is ambiguous. Take a child's ability for example, whilst this is likely to increase skills, it is not clear whether this would increase or decrease time in formal childcare. Whilst some parents may see their child's natural ability as a sign they will succeed in formal childcare at a young age, others may assume their child does not require the early intervention.

Furthermore, consistent estimates require the OLS model to be correctly specified. However, there is no consensus on whether the relationship between childcare and children's outcomes is linear or additive.

2.2.3 Cognitive and non-cognitive skills

The literature has examined the impact of attending childcare on both cognitive and non-cognitive development. Whilst cognitive skills are normally measured by test scores in a range of subjects, studies researching the impacts on non-cognitive outcomes have used a wide range of measures, such as behaviour, personality traits, goals, motivation, and preferences. Non-cognitive skills are generally seen as more adaptable for longer periods of time, compared to cognitive skills.

Research on the impact of formal childcare on cognitive skills has produced mostly positive evidence (Berger et al. 2021, Fort et al. 2020, Kuehnle & Oberfichtner 2020, Felfe & Lalive 2018, Carta & Rizzica 2018, Del Boca et al. 2016, 2018, Caprera 2015, Blanden et al. 2016, Gregg et al. 2005). For non-cognitive skills, this evidence is more mixed. Studies have found positive (Gupta & Simonsen 2010) but also zero (Kuehnle & Oberfichtner 2020, Hansen & Hawkes 2009) and negative (Fort et al. 2020, Baker et al. 2019, Magnuson & Duncan 2016, Burger 2010) effects.

Most of the earlier research relied on the selection on observables assumption holding and used matching methods to analyse the impact of childcare (see (Apps et al. 2013, Loeb et al. 2007, Magnuson et al. 2007, Goodman & Sianesi 2005) for UK and US specific research). Below we will discuss in more depth the research strategies and results of more recent studies.

Regional variation

The most common source of identification used within this literature is exploiting regional variation in childcare policy.

Berger et al. (2021), Felfe & Lalive (2018), Gupta & Simonsen (2016) and Felfe et al. (2015) all exploit a policy change that led to regional variation in childcare. They focus on different countries, France, Germany, Denmark and Spain respectively and use policies introduced in the early 1990s to 2005.

Difference-in-differences

Felfe & Lalive (2018) and Felfe et al. (2015) use the difference in the initial speed of the expansion of childcare across areas following a policy change. Their assumption is that the speed of the expansion does not relate to child development outcomes. Their analysis reveals no pre-policy differences in child development or socio-demographic factors between areas of fast- and slow-expansion. Also, the targeted level of early childcare is constant across both types of districts. They both use a difference-in-differences design, where they categorise fast-expansion districts as those that increased their centre-based care by more than the median expansion. The areas with expansion less than the median are the control group.

In Spain during the early 1990s there was an expansion of publicly subsidised full-time childcare for 3-year-olds, providing universal access to high-quality public care regardless of family background. Although the reform was national, the responsibility for implementing it was at a local level, resulting in a 10 year roll out that varied considerably across states. Felfe et al. (2015) uses this initial variation in the pace of the public childcare expansion to examine the impact of high-quality public care on children's educational outcomes. They specifically examine the long run impacts by examining the educational achievements of 15-

year-olds who were 3 years old before and after the reform in regions where public childcare saw substantial growth compared to those with a less pronounced increase.[4]

They find that the introduction of universal high-quality public childcare for 3-year-olds led to a 0.15 standard deviations increase on reading test scores at age 15. They find no significant effect of the reform on children's maths performance. Additionally, they find some evidence of a reduced likelihood of grade retention, with the probability of falling behind a grade in primary school decreasing by 2.5 percentage points.

Felfe & Lalive (2018) using a similar methodology, exploit the uneven expansion of childcare across German school districts when from 2005, German authorities channelled substantial funding into school districts to expand childcare.[5] In contrast to Felfe et al. (2015), Felfe & Lalive (2018) focus on the short run impact on non-cognitive outcomes. Their results indicate a positive effect of early childcare on children's motor and socio-emotional skills, at age 6, of 25 percentage points.

Instrumental variables

Berger et al. (2021) and Gupta & Simonsen (2016) also exploit regional variation to generate instrumental variables. Rather than having variation in the speed of implementation, these studies exploit the fact that only some regions experienced a policy change. Instrumental variables will correct for endogeneity if the instrument is correlated with enrolment in childcare and only impacts child outcomes through its effect on enrolment in childcare.

Gupta & Simonsen (2016), using data for Denmark, utilise the fact that only some munic-

[4]They order regions based on the percentage point increase in public childcare enrollment for 3-year-olds during the years following the reform, specifically from 1990 to 1993, and divide the sample at the median increase of 21%.

[5]They construct a difference-in-differences model in which fast-expansion districts are defined as those that increased their center-based care by more than the median expansion during the initial years following the reform, specifically for the 2011 and 2012 school entry cohorts. All other districts are categorised as part of the slow-expansion group.

ipalities provide guaranteed access to centre-based day care. They construct an instrument which takes the value one if the child lives in a municipality that provides guaranteed access to centre-based day care and is zero otherwise. Berger et al. (2021) uses data on children born in France in 2011, interacting spring births with local crèche availability—measured by the number of crèche slots per 100 children aged 3 and younger in the municipality—to create an instrumental variable. Their findings show that children born in the spring have a higher likelihood of securing a place in a crèche compared to those born at other times of the year, due to slots opening up when older children transition to preschool in September.

Gupta & Simonsen (2016) focus on the long run impact of childcare on cognitive skills. They estimate that being enrolled in centre-based day care at age 2 significantly increases GPA in Danish by 0.23 standard deviations as well as the probability of early academic track high school enrolment by 11 percentage points at age 16. Berger et al. (2021), focus on both the cognitive and non-cognitive short run impact of childcare and find that crèche attendance has a positive impact on language skills (0.34 SDs) and a negative impact on good behaviour (0.36 SDs). When assessing the impact across different types of non-parental care, children attending crèche scored highest, followed by those cared for by a private nanny. Children in parental care demonstrate the poorest language skill but were the best behaved.

Despite the difference in methodology used and country analysed, Gupta & Simonsen (2016) and Felfe et al. (2015), who both estimate the long run impact of childcare on cognitive skills, find very similar results for reading and writing subjects. Though, whilst Felfe et al. (2015) find no long run impact of childcare on maths, Gupta & Simonsen (2016) estimate a short run impact of 0.10 standard deviations.

The literature does not find similar findings for non-cognitive skills. Berger et al. (2021) estimate negative impacts whilst Felfe & Lalive (2018) find positive impacts.

The studies discussed here present results which are consistent with the whole literature.

There is some consensus that childcare has a positive impact on cognitive skills.[6] Regarding non-cognitive skills the findings are a lot more varied.

Focusing on non-cognitive skills, despite these studies using largely comparable empirical strategies, there are observed discrepancies in findings which may be driven partly by differences in country and therefore institutional context which highlights the lack of generalisability. There are also numerous ways to measure non-cognitive skills whilst the measure of cognitive skills is more generalised.

2.2.4 Heterogeneity

The range of findings within the literature could point to large heterogeneous impacts. Below we will discuss some of the heterogeneous factors found within the literature including, socio-economic background, gender, and quality of childcare.

Socio-economic background

The majority of the childcare policies introduced across the world have been focused on children from disadvantaged backgrounds. As a consequence, a large amount of the literature also examines the impact of childcare on children from disadvantaged backgrounds.

Three hypotheses are used to explain the relationship between socio-economic background and childcare. Firstly, the compensatory hypothesis suggest that childcare can be especially beneficial for low-SES children, offering cognitive stimulation and learning opportunities that may be lacking at home. On the other hand, the loss of resource hypothesis argues that children from middle- to high-SES families might miss out on the enriching stimulation

[6]Studies have not tended to find negative impacts on cognitive skills, but some have found insignificant impacts, for example, Baker et al. (2019), who exploit Quebec's childcare policy which aimed to provide childcare places to all children aged 0–4 in the province at a price of $5 per day.

and learning their families provide when placed in non-maternal care. Lastly, the dual risk hypothesis suggests that for children from already disadvantaged backgrounds, maternal employment may introduce an additional risk factor.

Providing evidence in support of the compensatory hypothesis, Geoffroy et al. (2010) show that children of mothers with low levels of education who attended formal childcare have school readiness and achievement test scores at ages 6 and 7 that are comparable to those of children with highly educated mothers. This finding is supported in a more recent study by Drange & Havnes (2019) in Norway, which shows that children from low-educated or low-income families who attended early childcare centers performed better in language and mathematics tests at age seven (by about 25% of a standard deviation), while the impact on children from high-income families was modest and not statistically significant. Similarly, Felfe & Lalive (2018) and Cornelissen et al. (2018), using data from Germany, found that early center-based care positively influences both cognitive and non-cognitive outcomes (such as language, motor, and socio-emotional skills) for children from disadvantaged backgrounds. Most recently, Carneiro et al. (2024) provided evidence that the Sure Start program in England was particularly effective for children from low-income backgrounds, with positive effects persisting at age 16.

The loss of resource hypothesis was investigated in the early research. Caughy et al. (1994) found that children from families with stimulating home environment who attended center-based childcare during infancy had lower mathematics scores. Gregg et al. (2005) report negative effects of maternal employment on children from families where mothers had high levels of education. More recently, Havnes & Mogstad (2015) find that the long run earnings benefits of childcare are driven by children from low-income households and that children of upper-class parents tend to experience an earnings decline.

There is evidence against the dual risk hypothesis as being a working mother does not necessarily imply reduced time with the child. Bianchi (2000) shows that, despite the increased amount of time US mothers spend away from home, the quality time and attention they devote to their children has remained relatively stable. Working mothers make an effort to maximise their time with their children. Additionally, higher parental educational attainment and work experience are positively correlated with the ability to educate young children. Jessen et al. (2021) demonstrate that childcare and parental investments complement each other rather than serve as substitutes, particularly for mothers with lower levels of education.

Finally, some studies find little evidence of an interaction between sociodemographic characteristics of the family and childcare. For example, Blanden et al. (2022) find no difference in the impact of an additional term spent in childcare for children from advantaged or disadvantaged backgrounds[7].

The difference in findings provides unclear answers on the impact of early childcare provision for decreasing early inequalities in child development.

Gender

Another characteristic across which studies find differing impacts is gender.

Havnes & Mogstad (2015) explore an expansion of childcare in Norway and report positive impacts of childcare on earnings at age 30-40. Felfe et al. (2015) exploit state-level variations in the expansion of childcare in Spain and find improvements in reading skills at age 15 of 0.15 standard deviations. Both these papers find their results are driven by females.

However, there is not an agreement on this finding as a recent study for Italy by Fort et al. (2020) estimates a negative effect of non-parental care on cognitive outcomes, particularly for girls, whilst Blanden et al. (2022) find no significant impact for females.

[7] Disadvantaged is measured by free school meal eligibility and deprivation in the neighbourhood of residence.

Quality

Another dimension that has been examined within the literature is quality. The quality of childcare has been shown to be a large determinant of the outcomes (Sammons et al. 2015). Measuring the quality of childcare is however sometimes a challenge. Within the literature, quality is normally determined by the author using institutional knowledge.

Varying quality could explain the difference between studies which find positive impacts and those that estimate negative impacts.

Denmark is known to have high quality childcare where every child is guaranteed a place, it is therefore not surprising that studies like Gupta & Simonsen (2010), find that children benefit more from centre-based care than from lower-quality informal day care. On the other hand, childcare in Québec are not guaranteed and therefore children experience varying quality. This quality difference could explain Baker et al. (2019) negative findings. Italy is another country where quality assessments of childcare are delivered at a local level which creates large variation. Fort et al. (2020) who use data from Italy, also find negative effects of non-parental care arrangements on cognitive development.

Grandparent care has been used as a comparison group in a number of studies. Whilst many suggest that a lower child to adult ratio benefits the child, they argue grandparent care is of very low quality due to the lack of structure. Bernal & Keane (2011), using US data, find formal centre-based care has no adverse effect, whilst informal care (grandparents, siblings, relatives, non-relatives), has a significant adverse effect. Hansen & Hawkes (2009) report similar findings for the UK, noting no impact of formal childcare at 9 months on child behavior at age 3. However, they observe that children cared for by grandparents experience more peer-related problems.

Focusing on the heterogenous impacts of childcare, further highlights the lack of generalisability within the literature. There is some consensus that higher quality childcare is beneficial for all. However, defining high quality is a challenge. Whilst some may see grandparent care as high quality due to the loving environment and low child to adult ratio, studies consistently find negative impacts of grandparent care. There is also some agreement that childcare is most beneficial for disadvantaged children however the impact for non-disadvantaged children is mixed.

2.2.5 Intensity of childcare

Whilst many childcare policies focus on the age funding for childcare starts or the hours that are subsidised, fewer studies have focused on the intensity of the treatment.

Intensity can be defined in several ways including the length of day, the number of days/hours attended per week, and the age that children start formal childcare. Within the literature, several aspects of intensity have been investigated. Peter et al. (2016), Kuehnle & Oberfichtner (2020), Kottelenberg & Lehrer (2014) and Blanden et al. (2022) all examine the impact of starting childcare at earlier ages on outcomes after starting school. Gupta & Simonsen (2010) and Berger et al. (2021) analyse the impact of hours spent in non-parental care.

If we think of early childcare as an investment in children's human capital, it is natural to think that the more time spent in childcare the better. Despite being theoretically straight forward, empirically this question has been difficult to answer.

Age childcare started

Peter et al. (2016), Kuehnle & Oberfichtner (2020), Fort et al. (2020), Blanden et al. (2022) and, Morando & Platt (2022) all examine the impact of starting childcare at a younger age on child outcomes. Peter et al. (2016) and Blanden et al. (2022) both use data from England whilst Morando & Platt (2022) focus on Ireland. Kuehnle & Oberfichtner (2020) and Fort et al. (2020) use data from Germany and Italy respectively.

Blanden et al. (2022) and Kuehnle & Oberfichtner (2020) both exploit variation in the month of birth to examine the impact of starting childcare earlier. Since 2004 all children have been entitled to free part-time early years education and care from the term after their third birthday. As a result, children become eligible at different times in the year. As all children in a school cohort begin formal schooling at the same time, regardless of their birth date, some will be eligible from more time in childcare.[8] Germany has a similar system where children are legally entitled to a place in childcare from their third birthday, though some start earlier. Most children start childcare in the summer of the calendar year in which they turn 3, creating a pattern where those born later in the year often start before their third birthday, while those born earlier start after. This results in a discontinuity in childcare starting age between December and January.[9]

They both exploit this discontinuity in a regression discontinuity design. Blanden et al. (2022) focus on children born around the December and March cutoffs described above. They estimate the impact of starting childcare 3.5 months earlier at age 3 on school achievement at age 5. They find that eligibility for an extra term of childcare leads to just under a 1 percentage point increase in the probability of reaching the expected level of competencies at age 5, compared to children not eligible for the extra term. Kuehnle & Oberfichtner (2020) find no short or long run effect of starting childcare earlier. They interpret this finding as the difference in quality of care in the home environment and that in childcare being too minor to matter.

[8]Children born between September 1st and December 31st can claim their free hours from the following January, those born between January 1st and March 31st from April, and children born between April 1st and August 31st from September of the next school year.

[9]Children born between September and December tend to start childcare before turning 3, with many born in December beginning in the summer before their third birthday. Children born in January typically start in the summer after their third birthday. In some areas, January-born children are restricted from starting childcare before they turn 3, further contributing to the age discontinuity between December and January.

Fort et al. (2020) also exploit a discontinuity in their analysis by examining admission thresholds. In Bologna, day care applicants rank their preferences for available programs and are assigned to priority groups based on observable family characteristics. Within each priority group, applicants are ranked according to a family affluence index (FAI), which adjusts for household size and combines family income and wealth. The FAI thresholds for available spots are set each year; applicants whose FAI meets or is below the threshold for their top-choice program receive an offer, while those with a higher FAI may be admitted to a less preferred program or excluded altogether. They focus on the discontinuity created by whether a child received a place at their first-choice program or not. To assess the impact of starting age in childcare, they examine how this discontinuity interacts with whether the application was the child's first attempt.

Exploiting these discontinuities, they find that one additional day care month at age 0–2 reduces intelligence quotient by 0.5% (4.7% of a standard deviation) at age 8–14 in a relatively affluent population. For non-cognitive skills, an extra month decreases openness and agreeableness by 1.4% and 1.2% and increases neuroticism by 0.9%.

Peter et al. (2016) and Morando & Platt (2022) do not adopt a quasi-experimental design, instead they make use of the large amount of data they have access to. Peter et al. (2016) utilise the data on individual and parental characteristics and estimate their effects using propensity score matching. Morando & Platt (2022) employs a cumulative value-added approach to integrate measures of childcare type at both 3 years and 9 months of age, while also controlling for skill measures at those same ages.

Peter et al. (2016) estimate the impact of starting childcare after 2.5 years old compared with starting earlier on non-cognitive skills at age 5 and 7. Morando & Platt (2022) compare center based care with other forms of non-parental care at ages 9 months and 3 years, to consider the impact of childcare type on three domains of socio-emotional skills: externalising, internalising and prosocial behaviours.

Peter et al. (2016) find evidence to suggest that starting childcare later increases peer problems and reduces prosocial behaviour. These findings are significant at age 7 but not age 5. At age 7, they estimate a 6.7% standard deviation increase in behavioural problems if they enter childcare later than age 2.5. Morando & Platt (2022) find that starting childcare at 9 months is associated negatively with both externalizing (0.18 SD) and prosocial behaviour (0.11 SD), measured at age 5, versus not being exposed at all. They conclude however, that any effects at 9 months work through continued exposure up to and including 3 years.

Overall, this area of literature reinforces the findings above that childcare has a positive impact on cognitive skills which is stronger the earlier the child starts childcare. For non-cognitive skills the picture is mixed. Whilst there is some evidence that childcare increases non-cognitive skills with increased benefit the earlier childcare is started (Peter et al. 2016), some also find the oppostie (Fort et al. 2020, Morando & Platt 2022).

Hours in childcare

The psychology literature has numerous studies that have compared full-day with half-day kindergarten (Cryan et al. 1992, Plucker et al. 2004, Votruba-Drzal et al. 2008, Zvoch 2009, Pelletier & Fesseha 2019), with most finding positive impacts of full-day sessions. Whilst providing interesting findings, they all focus on the US.

Gupta & Simonsen (2010) and Berger et al. (2021) are two of very few studies to evaluate the effect of hours in non-parental care on child outcomes outside of a US context. The methodologies adopted by these studies are discussed above, here we will describe how they expand their analyses to focus on the intensity of childcare.

Gupta & Simonsen (2010) use data from Denmark of individuals born in 1995, whilst Berger et al. (2021), use French data on children born in 2011. They both examine the impact of hours in childcare on cognitive and non-cognitive skills though Gupta & Simonsen (2010) focus on the long run impact whilst Berger et al. (2021) only consider the short run.

Gupta & Simonsen (2010) examine the marginal effect on outcomes of small increases in the hours spent in non-parental care. By focusing on local comparisons, they minimise the likelihood of indirect effects and provide a clearer interpretation of the direct impact. Additionally, their estimator accounts for the possibility that selection into non-parental care may depend on unobservable factors, but once the decision to use non-parental care is made, the choice of hours is assumed to be based on observable factors.

They find that increases in the hours from 0-10 to 10-20 and 10-20 to 20-30 are benign, no matter the choice of care. This is unsurprising as under 30 hours in formal childcare allows for a good mix of time with parents and time with peers. Further increasing hours, however, seems to significantly worsen non-cognitive skills, and this is significant in the case of preschool.

Berger et al. (2021) results suggest that additional hours or days per week in crèche are linked with improved language and motor skills. Compared to children in other types of care, those attending crèche for three, four, or five days per week show language skills that are 0.10, 0.19, and 0.20 standard deviations higher, respectively. For motor skills, children who attend crèche five days per week exhibit skills that are 0.15 standard deviations better than those in other care settings. Conversely, children spending three or five days a week in crèche demonstrate 0.12 and 0.10 standard deviations poorer behavior compared to their peers in other types of care. It should be noted that these estimates are taken from an OLS estimation and therefore are likely biased.

Morris et al. (2021A), whose research is part of the Study of Early Education and Development (SEED), closely aligns with the focus of this research. Using data from England, Morris et al. (2021A) explores the impact of a policy change, when in 2013/14 free early education was introduced for disadvantaged two-year-olds, on their cognitive and non-cognitive skills at age 3. Using a regression discontinuity design based on birth dates, similarly to Blanden et al. (2022) as discussed above, they compare children with different eligibility but born within a short period, assuming that other factors affecting childcare use were equivalent due to their close birth dates. Their findings suggest that use of formal group childcare is linked to higher verbal ability, more prosocial behaviour and fewer emotional symptoms and peer problems. However, it is also associated with poorer socio-emotional outcomes in conduct problems and emotional self-regulation. Subgroup analysis reveals that these poorer outcomes are only present among children attending more than 35 hours per week of childcare.[10]

Intensity of childcare has received less focus in the literature making it difficult to reach a conclusion. This area of the literature further highlights the variation in findings. As this area has huge policy focus, there is a need to build this area of research.

2.2.6 Critical analysis

Despite methodological improvements, weakness still exist.

Omitted variables

Estimating the causal effect of childcare is challenging due to selection bias, which occurs when the factors influencing child development are correlated with unobserved characteristics of parents and children that also affect the decision to use childcare. Two main sources of selection bias that are particularly concerning: (1) parents, especially mothers who choose to place their child in childcare, may differ systematically from those who do not; (2) a child's

[10]Melhuish & Gardiner (2020) provides a follow up to this research when the children were aged five. Melhuish & Gardiner (2020) finds negative impacts of time in childcare on non-cognitive skills at age 5. The variation in findings between age 3 and 5 is not uncommon in the literature and something we discuss in more detail further on. Research has also been carried out at age 7 but only focused on cognitive skills.

cognitive and non-cognitive abilities may influence a mother's decision to use childcare. For instance, if highly educated mothers are more likely to have children with high cognitive and non-cognitive abilities and also more likely to use childcare, statistical analyses might incorrectly attribute any observed positive effects on child outcomes to childcare use rather than to these unobserved factors.

Two key factors discussed by Bernal & Keane (2010), which are often overlooked in the literature, are maternal time input and goods inputs. Both variables are not directly observed. Most studies ignore this problem, simply using maternal employment or childcare use in place of maternal time. What they do, however, with the time spent with their children matters. For example, reading with them, taking them to museums or just placing them in front on the television. Many studies proxy good inputs with household income, however having the ability to provide does not always translate into actually providing.

Many studies use administrative data with limited information on background characteristics. For example, Blanden et al. (2022) have no information on family background and use free school meal eligibility as a proxy for household income which has two main limitations. Firstly, it may not identify some groups, such as children in 'working poor' households (Hobbs & Vignoles 2010). Additionally, free school meal eligibility is linked to parental employment which can change through a pupils time at school. Other measures of disadvantage, for example low parental education, are persistent (Ilie et al. 2017).

Studies using large datasets, specifically collected on preschool-age children, have access to in depth information on parents. Research such as Sylva et al. (2004) and Vandell et al. (2010), find that reading with the child, teaching songs and nursery rhymes, painting and drawing, playing with letters and numbers, visiting the library, teaching the alphabet and numbers, taking children on visits and creating regular opportunities for them to play with their friends at home, were all associated with higher intellectual and social/behavioural scores. Household income and parents' education may not fully account for the time factor, as Sylva et al. (2004) demonstrated that the home learning environment was only moderately correlated with parents' educational or occupational levels. Instead, it was more strongly related to children's intellectual and social development. In other words, the activities and interactions parents have with their children are more significant for their development than the parents' own educational or occupational status.

Generalisability of findings

To justify expanding universal early childcare programs, policymakers often refer to the positive effects found in the literature. In reality, there is limited consensus. Depending on the country, time frame and methodology used, the estimated impact can vary considerably. The range of findings found highlights the lack of generalisation.

As the direction of the bias between childcare and child outcomes is ambiguous, it is difficult to make a judgement as to what size we expect the estimates to be. Many studies present results which are larger than their initial naive estimates, (Berger et al. 2021, Felfe & Lalive 2018, Gupta & Simonsen 2016) some studies find positive average effects (Drange & Havnes 2019, Carta & Rizzica 2018, Gregg et al. 2005, Del Boca et al. 2016, Caprera 2015, Felfe & Lalive 2018, Blanden et al. 2016, Kuehnle & Oberfichtner 2020, Berger et al. 2021) and others documenting no effects (Carta & Rizzica 2018, Hansen & Hawkes 2009, Kuehnle & Oberfichtner 2020) or even negative effects (Magnuson & Duncan 2016, Baker et al. 2019, Burger 2010, Fort et al. 2020).

The mixed results may partly stem from the varied country contexts studied. Most research on the relationship between childcare and child outcomes is concentrated in the US and the UK, with fewer studies examining data from Europe and other regions. Kamerman & Waldfogel (2005) suggest that differences in findings are due to significant variations in access to and quality of childcare. English-speaking countries often rely on market-based childcare systems, while European countries tend to have more regulated, homogeneous, and universal childcare services (Spiess et al. 2003). As mentioned above, these quality differences could be used to explain the difference in findings from Canada (Baker et al. 2019) and France (Berger et al. 2021).

Andrew et al. (2021) suggest that the different shares of non-working mothers across countries are related to the difference in outcomes. With more than 85% of mothers employed, the highest rates are found in Iceland, Slovenia and Sweden. The United Kingdom falls just below 75%, whilst for the United States, 70% of mothers are employed. At the other end of the spectrum, the rate is below 60% in 6 OECD countries (Chile, Greece, Italy, Korea, Mexico, and Turkey), with particularly low rates for Mexico (47%) and Turkey (30%) (OECD 2023). As working mothers are more likely to place their children in childcare and with more policies focused on working parents, the characteristics of the sample examined across these countries could differ. As mentioned above, the loss of resource and the dual risk hypotheses would suggest different outcomes depending on the groups analysed.

Also, much of the literature focuses on children aged 3–5 as this is when the majority of children start childcare. Recently, there has been policy focus on childcare at ages younger than 3. There is also a theoretical suggestion that it is in infancy that children may be most susceptible to influences on their development (Magnuson & Duncan 2016).

Some of the literature find that results fade out quickly and are no longer significant by age 7 (Blanden et al. 2022, Deming 2009, Elango et al. 2015, Garces et al. 2002). Felfe et al. (2015), however, find positive effects at age 15. Gupta & Simonsen (2016) and Peter et al. (2016), who both examine the impact of childcare for children who start younger than 3, find persistent impacts.

Focusing on non-cognitive development, both Felfe & Lalive (2018) and Berger et al. (2021) focus on children younger than 3 but find differing results. Felfe & Lalive (2018) estimate positive impacts on children's motor and socio-emotional skills whilst Berger et al. (2021) find insignificant impacts on motor skills and negative effects on behaviour.

There is some suggestion that starting childcare earlier could lead to more persistent impacts however, due to the lack of research that focuses on younger aged children, it is difficult to make a valid conclusion.

Overall, the variation in findings indicates the need for continued research in this field.

Measure of outcomes

Despite research questions being similar across studies, the way the dependent variables are measured varies considerably.

For studies using cognitive development, they focus on a range of measures with some using national tests whilst other utilise teacher assessments. Even when examining the impact on a subject such as maths which tends not to vary across countries, studies find differing results. Felfe et al. (2015) found no impact of childcare on maths, whilst Gupta & Simonsen (2016) estimated an impact of 0.10 standard deviations.

Non-cognitive skills are even more varied across studies as they are multidimensional. A broad array of traits has been identified as non-cognitive skills, encompassing personality and behavioral attributes such as persistence, motivation, temperament, attention, communication, confidence, and self-esteem. Specifically at younger ages, non-cognitive skills are normally reported by parents or teachers. One commonly stated disadvantage, therefore, is the concern that the reporters are able to distort their responses to create an inaccurate impression. Mothers may feel guilty for not spending a lot of time with their child and therefore could inflate their responses to make their child seem better behaved.

Morris et al. (2021B) investigate the consistency of non-cognitive skills and find evidence suggesting that measures intended to capture these skills may suffer from significant response variability or measurement error. They report low correlations between non-cognitive measures over time.

They suggest that the SDQ scale might be the most reliable tool, as it includes a comprehensive set of validated questions rather than relying on impulsivity measures derived from direct child behavior assessments. Additionally, they found that results using aggregated non-cognitive skill values were more accurate than those based on individual measures. Aggregating data from multiple occasions helps reduce measurement error compared to single-instance measures.

2.2.7 Research Gap

Analysing the relationship between childcare and child outcomes is impacted by methodological challenges. Research has varied considerably in terms of methodologies, time spans, countries and outcomes measured, generating conflicting findings. In 15 studies covering polices spanning over 20 years, research has broadly found that childcare has positive effects on cognitive outcomes, children who start childcare at a younger age benefit more, however the effects could be short lived. On the other hand, the current research has failed to reach an agreement on the impacts of childcare on non-cognitive skills. The impact of childcare on non-cognitive skills in comparison to cognitive skills seems to be more impacted by the county analysed, specific measure of skills, starting age and quality of childcare. This overview points towards important heterogeneity in the effects of early childhood intervention, specifically on non-cognitive skills, highlighting the need for further research.

This research, therefore, aims to make three contributions. Firstly, add to the knowledge base on the impacts of childcare. Due to the variation in findings, providing useful evidence for policy guidance is a challenge, by increasing the evidence we might start to be able to draw some conclusions. Secondly, this research specifically focuses on the impact of hours in childcare. The impact of the intensity of childcare is an area of the literature which has received less attention. Finally, we focus on childcare younger than 3, while the majority of the literature has examined the impact of childcare on children aged 3-5. As Baker (2011) states in his survey of the literature "the case for universal early childhood interventions does not have a strong foundation in evidence", suggesting we need to identify if there is an optimal age for children to begin attending childcare.

2.3 Data

This research uses data from the Millennium Cohort Study (MCS). The MCS is a multidisciplinary cohort survey run by the Centre for Longitudinal Studies at the University College London. The study is a valuable data source as it tracks the lives of a sample of about 19,000 babies born in the UK between 1st September 2000 and 11th January 2002.[11] The sample was constructed to be representative of the total UK population. The data collectors selected electoral wards with the aim to recruit 100 per cent of the children born in the eligible period within them.[12] They also wanted to adequately represent disadvantaged and ethnic minority children. The population of wards was therefore stratified by ethnicity and the Child Poverty Index. This means that certain sub-groups of the population were intentionally over sampled.

The survey is conducted in several waves, the first occurred when children were aged nine months, gathering information from the parents of 18,818 children. Since then, families have been interviewed again six times at ages 3, 5, 7, 11, 14 and 17.[13] The survey was originally

[11] 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).

[12] They achieved a response rate of 72 per cent of all the families with eligible children living at nine months in the sampled wards.

[13] Wave 8 at age 22, has been undertaken but the data has not been released.

answered by just the parents. From wave 2 onwards, the child was also surveyed along with older siblings. Class teachers responded to a survey in waves 3-5. Early topics included parental and child health, parenting activities and attitudes, physical, social and cognitive development of the child, preschool experiences, and leisure activities etc. As the cohort member got older, there was a larger focus on schooling, relationships, mental and physical health, wellbeing and, aspirations and attitudes. Many topics were covered in each wave including family composition, housing and local area, parental education, employment and income.

The main advantage of the MCS is the rich range of information regarding the experiences and outcomes of the MCS children and their families. This allows many individual and family characteristics to be controlled for. They also ask the same questions at multiple ages which allows for changes over time to be analysed. The main limitation of the survey is that the longitudinal pattern of response is complex, with attrition and re-entry. By age 17, 10,757 cohort members responded to the survey, a reduction of 8,061 from the original sample. Plewis (2007) argues that although the cases lost from the sample were different from those that remained, they were not substantially different. However, the reduction in the sample size impacts analysis particularly when examining sub-groups.

2.3.1 Sample

In this research we use data from the first six waves, which took place when the children were aged 9 months, 3, 5, 7, 11 and 14 years old.[14] We are interested in the impact of childcare by age 3, therefore the majority of our focus is on the information provided in wave 2. We use the data provided in wave 1, at 9 months, to construct our control variables and the information in waves 3-6 to measure non-cognitive skills at older ages.

[14]This research stops at age 14 as this is the last year mothers completed the Strengths and Difficulties Questionnaire, the main outcome of interest.

We restrict our sample to children living in England, so that the childcare supply is similar for all children observed. We also exclude twins and triplets since childcare arrangements and their effects may be different when more children need to be looked after. In the MCS parents are categorised as main parent and second parent. We focus on responses from the main parent.

The Strength and Difficulties questionnaire, the outcome of interest, as well as reported hours in childcare, suffer from non-responses, we keep cohort members who have responded at least once to the SDQ and report hours in childcare by age 3.[15] The initial sample is therefore 6,387 cohort members.[16]

2.3.2 Childcare measure

This research is interested in the impact of time spent in formal childcare by the age of 3. We define formal childcare as any type of childcare provided by someone who is not family or friend. This includes centre-based care for example nursery but also childminders. We do not include nannies or au pairs as we are interested in the impact of being cared for in a group setting outside the home. Nurseries are likely to be more structured than childminders however, both are required to have some level of childcare qualifications. We also believe that the difference in structure is less important for non-cognitive skills than cognitive development.

In the first three waves of the survey, the main parent was asked about the childcare choices made. In wave 1, when the child was 9 months old, working mothers were asked to state the types of care being used at the time of the survey when they were at work. In wave 2, when

[15]We examine the determinants of non-response as well as examine the impact of a change in the sample in Section 2.5.

[16]Focusing on England reduces the sample to 11,336. Removing twins reduces the sample by 316. The sample is reduced to 8,650 as there is large non-response to the Strength and Difficulties questionnaire. The final reduction in the sample is due to non-response to other variables included in the model.

the child was 3 years old, all mothers were asked details about the type of childcare that had been used since the first survey, including starting dates, stop dates, and the number of hours per week.[17] [18]

This research focuses on information from the second wave for two reasons. Firstly, it allows us to include working and non-working mothers, generating a more representative sample. Secondly, information on hours in childcare, the main focus of this research, is only surveyed in wave 2.

There is potentially measurement error in the number of hours of care reported by the main parent. Whilst some parents will know exactly how many hours they enrol their child in childcare for, others may only know how many sessions a week they take them. This uncertainty could lead to parents guessing how many hours this equates to. Whether the guessing or quick in head calculations would give under or over estimates is not clear.

2.3.3 Measure of non-cognitive skill

The outcome of interest in this research is non-cognitive development. We focus on non-cognitive skills for two reasons. Firstly, in comparison to cognitive skills, the literature has found more varied results meaning there is a lack of a consensus on the impact of childcare on non-cognitive skills. Secondly, we are interested in both the short and long run effect of time in childcare and it has been argued that non-cognitive skills are more malleable as well as more predictive of long-term outcomes than are test scores (Heckman & Rubinstein 2001, Lindqvist & Vestman 2011, Mueller & Plug 2006).

As our measure of non-cognitive skills we use the Strength and Difficulties Questionnaire

[17] In the third wave, when the child is 5 years old, all mothers are asked about all childcare prior to starting school. This research focuses on children under the age of 3 and therefore, focuses on the first 2 waves of the survey.

[18] Some 3 year olds were eligible for free childcare prior to 2004. In Table A.7 we reduce the sample to children who paid for all their hours of childcare. Our findings are robust to the change in the sample.

(SDQ), proposed by Goodman (1997), which screens the behaviour of 2–17-year-olds. The questionnaire is completed by the main parents and consists of 25 questions over five separate dimensions: Conduct Problems, Hyperactivity/Inattention, Emotional Symptoms, Peer Relationship Problems, and Prosocial behaviour. The questionnaire consists of statements to which the responses are: 'not true', 'somewhat true', and 'certainly true'. [19] [20] Higher scores on the first four dimensions and lower scores on the pro-social subscale indicate greater problems. A total difficulties score is generated by summing the first four scales and excluding the prosocial scale, which can be used as a positive counter measure to the overall SDQ score. The total difficulties score ranges from 0 to 40, the subscales from 0 to 10. For the total difficulties score, 0–15 is defined as having low needs, 16–19 some needs and 20–40, high needs. For our analysis we use the overall SDQ score as well as its dimensions to cover various aspects of a child's non-cognitive skills. In the estimations, the SDQ and all subscales are standardised to have a mean of zero and a standard deviation of one. We concentrate on the overall SDQ measure, as Morris et al. (2021B) suggest that using aggregated non-cognitive skill measures yields more precise results compared to individual measures. Aggregated measures, which incorporate data from multiple points, are expected to have less measurement error than individual assessments.

[19] Statements include: Often has temper tantrums or hot tempers, generally obedient, often fights with other children, often lies or cheats, steals from home, school or elsewhere, restless and overactive, constantly fidgeting or squirming, easily distracted and concentration wanders, thinks things out before acting, sees tasks through to the end, often complains of headaches, many worries, often unhappy and downhearted, nervous or clingy in new situations, many fears and easily scared, rather solitary and tends to play alone, has at least one good friend, generally liked by other children, picked on or bullied by other children, gets on better with adults than with other children, considerate of other people's feelings, shares readily with other children, helpful if someone is hurt, kind to younger children, often volunteers to help others.

[20] For negative statements 'Not true' is coded as zero, 'somewhat true' is given one point and 'Certainly true' two points. For positive statements, the opposite is true.

Figure 2.1: Distribution of total difficutiles at age 3

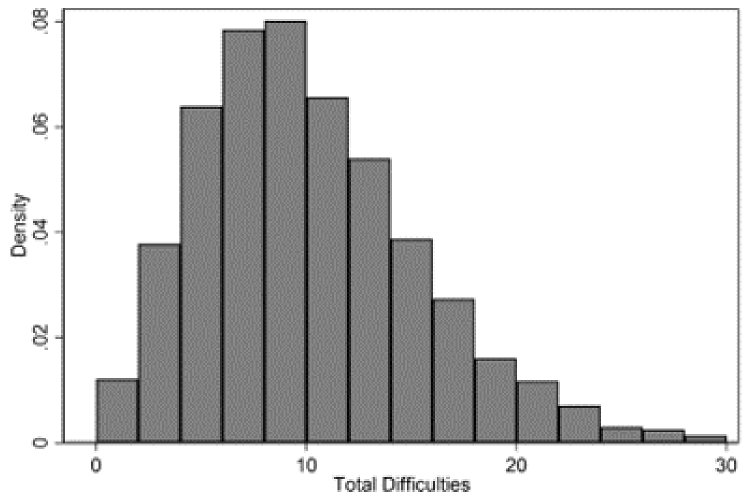


Figure 2.1 illustrates the distribution of the total difficulties at age 3. The distribution is normal with a few children reported to have very low or very high difficulties.

2.3.4 Conditioning variables

Cohort members were first surveyed at 9 months old, before many of them started childcare, and we therefore use the information given at age 9 month to construct the control variables. Due to the large amount of information held on both the children and their families we are able to control for many variables (see Table 2.2 in Section 2.5). We consider the child's characteristics (gender, ethnicity, birth-month, birth-weight, special educational needs and three indicators of child development[21]); household characteristics (number of siblings, single parent, real weekly equivalent income[22] and region of residence); and the main parent's

[21]We include three factors suggested by Del Boca et al. (2018). They are: "s/he waves bye-bye on her/his own when someone leaves", "s/he can pick up a small object using forefinger and thumb only", "s/he can sit up without being supported"; answers are "often", "once or twice", and "not yet".

[22]The equivalent income is the income of the household taking into account the number of people in the family and assigning weights. The one provided in the MCS follows the OECD equivalence scale, which assigns a value of 1 to the first household member, of 0.7 to each additional adult, and of 0.5 to each child.

characteristics (age and level of education[23] economic activity status and health).

2.4 Method

The data allow this research to observe hours spent in childcare by the age of 3 as well as non-cognitive outcomes from age 3-14. This analysis aims to identify the causal effect of hours spent in childcare by the age of 3 on non-cognitive outcomes both in the short run and the medium run. As previously discussed in Section 2.2, establishing causality between hours spent in childcare and non-cognitive outcomes faces methodological challenges. There are two primary threats to causality: (i) omitted variable bias, and (ii) measurement error. We discussed measurement error is Section 2.3.3. This section will outline how this research will try to correct for omitted variable bias.

The analysis starts by estimating an ordinary least squares (OLS) regression in which we regress total difficulties reported in each year on the average weekly hours spent in childcare by age 3 and the covariates.[24] The models take the following form:

$$TotalDifficulties_{it} = \beta_0 + \beta_1 Hours_i + \beta_2 Child_i + \beta_3 Family_i + \beta_4 Mother_i + \epsilon_{it} \tag{2.1}$$

where $TotalDiffculties_{it}$ is the total number of difficulties reported by the main parent in the strength and difficulties questionnaire for child i at age t. Hours is the average number of reported hours per week spent in formal childcare before age 3. *Child* is a vector of individual characteristics including gender, ethnicity, birth-month, birth-weight, special

[23]The level of education is controlled for by including a range of dummy variables relating to the level of qualification.

[24]We make the assumption that total reported difficulties depends linearly to hours in formal childcare. This assumption is tested in Section 2.5.3

educational needs and three indicators of child development. $Family$ is a vector of household characteristics including a dummy variable equal to one if they have older siblings, real weekly equivalent income, region of residence and presence of the father at home. $Mother$ is a vector of the mother's characteristics including age at birth, level of education, economic activity status and health.[25] ϵ_{it} captures the unobservable determinants of non-cognitive skills.

The main interest of this research is the estimation of β_1, which is the effect of hours in formal childcare on non-cognitive skills. To interpret β_1 as the causal effect of an hour in formal childcare by age 3 on an individual's non-cognitive skills, we require independent variation in hours in formal childcare, meaning the zero conditional mean assumption must hold, $E(\epsilon_{it}|Child_i, Family_i, Mother_i) = 0$. Due to the endogeneity of hours in formal childcare, it can be argued that this assumption may not hold.

Parents who enrol their children in formal childcare for more hours may differ in both observable and unobservable ways from those who opt for a few hours or none at all. These differences could be driven by child and parent factors. Parents of children with higher natural ability may place their child in formal childcare for more hours because they believe they will be able to cope with the separation from their parents better. On the other hand, they might enrol their child in formal childcare for fewer hours, due to a belief their child does not require early intervention. Parents with more time to devote to their child may have chosen fewer hours of childcare though having the time does not always lead to effective use. More "education oriented" parents may choose to place their child in childcare for many hours at an early age. They may also believe that they can provide their child with a strong early education at home and therefore place them in formal childcare for fewer hours. Many of these factors may be not observable and may influence the child's non-cognitive development, biasing the estimates. The direction of the bias is not straightforward. Take a child's ability for example, whilst this is likely to increase non-cognitive skills, it is not clear whether this

[25] All control variables are measured at age 9 months.

would increase or decrease hours in formal childcare. The same can be said for a number of potential omitted variables. The number of recorded hours in childcare may also be subject to measurement error as discussed in Section 2.3.3. As discussed in Section 2.2, much of the literature to date has suggested a negative bias, we therefore interpret the OLS results as a lower bound of the association between hours in childcare and non-cognitive skills.

2.4.1 Instrumental variables

To identify the causal effect of hours in formal childcare on non-cognitive skills, we exploit endogenous variation in hours by using instrumental variables.

The job characteristics of the parents, specifically the mother, are likely to be large determinants of the amount of time children are enrolled in childcare. Using the Labour Force Survey, which is a nationally representative study of the employment circumstances of the UK population, covering approximately 40,000 responding UK households per quarter, we obtain information on some job characteristics. We focus on two characteristics. Firstly, shift work. We make the assumption that mothers who work shifts are likely to place their child in formal childcare for fewer hours per week. This is due to the structured opening hours in formal childcare not working for the unsocial working hours of shift workers. The second characteristic we focus on is having variation in working hours each week. We believe that the uncertainty of hours of working will drive parents to enrol their child in formal childcare for more hours per week, in order to cover all possible times when they might be working.[26]

To try to increase the validity of the instruments we use the probability of shift work and uncertain hours calculated at the three-digit SOC code level. Using the Labour Force Survey, we calculate the percentage of individuals within a three-digit SOC code that report working

[26]These assumptions are tested in Section 2.5.

shift work/having varying hours which we link to the MCS using the three-digit SOC code in which the mother was employed whilst she was pregnant.[27] We use the job the mother was in during pregnancy as this can be argued to be pre-determined.[28] Further discussion on the quality of the instrument is provided in Section 2.5.

The average percentage of individuals within a three-digit SOC code that reported working shifts was 22.1% which relates to SOC code 221, health professionals for humans and animals. The occupation that reported the highest percentage of shift work at 75.3% was transport associate professionals. Workers in this group command and navigate aircraft and vessels. The occupation reporting the lowest percentage of shift work at 1.1% was architects, town planners and surveyors.

For uncertain hours of work, the mean was 45.5% associated with SOC code 213, information technology and telecommunication professionals. The SOC code associated with the highest percentage of worker with uncertain hours (80.5%) was 121, managers and proprietors in agriculture related services who plan, organise, direct, and control the activities and resources of agricultural, forestry, fishing and similar establishments and services. The SOC code associated with the lowest percentage of workers with uncertain hours (22.1%) was 923, elementary cleaning occupations.

We use a two-stage least-squares (2SLS) regression approach to first estimate the hours in formal childcare as a function of the probability of shift work and uncertain hours for the mother, net of child, household, and mother characteristics. The predicted hours in formal childcare is then forwarded to a second-stage regression to predict the unbiased local average treatment effect (LATE) of hours in formal childcare on non-cognitive skills. The first stage equation takes the following form:

[27] We also calculate the instrument at the one and two-digit SOC code levels. Results using these alternative instruments are shown in Table A.2 and A.3 in the appendix.

[28] The current occupation of the mother could be a function of childcare use, and so does not satisfy the requirements of an instrumental variable.

$$Hours_i = \lambda_0 + \lambda_1 Shift_i + \lambda_2 Uncertain_i + \lambda_3 Child_i + \lambda_4 Family_i + \lambda_5 Mother_i + \gamma_{it} \quad (2.2)$$

Where $Shift_i$ is the percentage of individuals, working in the same three-digit SOC code as the mother at the time of pregnancy, who reported working shift work, and $Uncertain_i$ is the percentage of individuals working in the same three-digit SOC code as the mother at the time of pregnancy, who reported having variation in their weekly hours.

The second-stage equation takes the following form:

$$TotalDifficulties_{it} = \beta_0 + \beta_1 H\hat{ours}_i + \beta_2 Child_i + \beta_3 Family_i + \beta_4 Mother_i + \epsilon_{it} \quad (2.3)$$

where $H\hat{ours}_i$ is the predicted hours of childcare based on the first stage.

The IV strategy requires that three assumptions be met. First, the instruments must be relevant, meaning that they are highly predictive of hours in formal childcare. Second, they must be exogenous, meaning they are not correlated with the error term in the explanatory (second-stage) equation. Finally, the instruments affect non-cognitive skills only through their effect on hours in formal childcare and not through any other pathway. We discuss the second and third assumption together.

The first assumption is easily tested and, as shown in the first stage results (Table A.4), holds true in all our models. The second may be violated if, for example, mothers select into professions based on children's needs.

Although we cannot fully rule out this possibility, we attempt to minimise it by arguing that mother's occupation at the time of pregnancy is predetermined. We also adjust for an extensive array of covariates. We control for characteristics of the mother including education level, age and health as well as household income, presence of the father and a first-born indicator. Adjusting for these factors should reduce the risk of bias. In the robustness checks we test this assumption further by examining why they left their previous job.

2.5 Results

2.5.1 Descriptive Statistics

This research focuses on hours in formal childcare for children under the age of 3. We focus on an initial sample of 6,387 individuals who record at least one year of non-cognitive skills as well as hours in childcare (see Section 2.3 for further discussion on construction of the sample.) Within the sample, 13% of children have attended formal childcare by the age of 3.[29] 87% have not yet attended formal childcare, of whom 83% have only been cared for by their parents, leaving 17% who have received informal childcare, mainly from the maternal grandmother.

The main variable of interest within this research is non-cognitive development, measured by the parent's response to the strength and difficulties questionnaire as discussed above. Table 2.1 presents the average number of difficulties reported for ages 3, 5, 7, 11 and 14.[30] Children who receive formal childcare by age 3 have statistically significantly lower reported difficulties at every age compared to children who receive no formal childcare by age 3.[31]

[29] Blanden et al. (2016) shows that in the year 2000 most three year olds (82%) were already attending some type of pre-school education. Huskinson et al. (2013) shows that in 2011, 63% of 0-4 year olds were using formal provision with this statistic being driven by the increase in uptake of formal childcare from age 3.

[30] Standard deviation is in parentheses.

[31] Table A.1 in the appendix presents total difficulties at age 3 for different types of childcare

Table 2.1: Average number of reported difficulties

	Childcare	No childcare
Total difficulties age 3	8.01 (4.30)	9.96 (5.33)
Total difficulties age 5	6.16 (4.08)	7.37 (4.98)
Total difficulties age 7	6.36 (4.64)	7.55 (5.41)
Total difficulties age 11	6.83 (5.44)	7.78 (5.82)
Total difficulties age 14	7.08 (5.69)	8.12 (5.89)

Descriptive statistics for the covariates, shown in Table 2.2[32], reinforce that there is likely systematic selection into childcare. Focusing on individual characteristics, children of white ethnicity, children born with a higher birth weight, having no learning difficulties and having fewer siblings are all associated with a greater likelihood of attending childcare by age 3. Differences are also evident for family and parental characteristics. For example, children of less educated, lower-income, and younger parents are less likely to attend formal childcare by the age of 3.[33] All differences in the covariates, apart from gender and development measures one and two, are statistically significant.

[32]Standard deviation is in parentheses.

[33]Mother's education is a dummy variable equal to one if the mother has a degree or equivalent. The household income is the income of the household taking into account the number of people in the family and assigning weights. We also control for birth month and region of residence, the sample is evenly distributed across these measures.

Table 2.2: Summary statistics for covariates

Variable	Childcare	No Childcare
Female	48.02% (0.50)	49.41% (0.50)
White	90.29% (0.30)	82.02% (0.38)
SEN	6.11% (0.24)	9.20% (0.29)
Birth Weight	3.42 (0.51)	3.35 (0.58)
Development 1	1.13 (0.42)	1.14 (0.44)
Development 2	1.97 (0.83)	1.94 (0.83)
Development 3	1.02 (0.17)	1.05 (0.28)
Number of Siblings	0.56 (0.72)	0.94 (1.03)
Year of mother's birth	1969 (4.98)	1972 (5.85)
Mother degree	49.22% (0.50)	18.44% (0.39)
Mother employed	86.95% (0.34)	54.12% (0.49)
Equivalised weekly household income	529.82 (250.08)	322.93 (216.39)
Single parent	9.58% (0.29)	13.31% (0.34)
Mother good health	88.02% (0.32)	83.32% (0.37)

This research is predominantly interested in the impact of the intensity of childcare measured by hours. Conditional on being cared for by someone other than their parents, the average number of hours per week being cared for by someone other than their parents is 20. Table 2.3 shows that children who attend formal childcare by age 3 on average, spent 27 hours per week in that setting whilst children in informal care settings spent 17 hours per week.

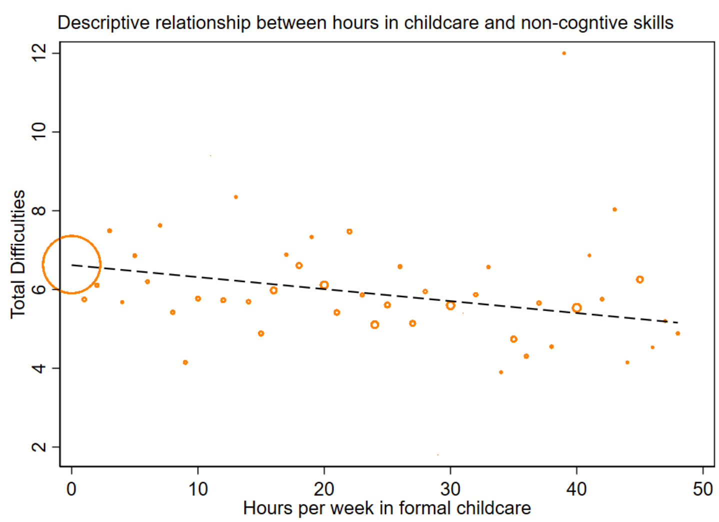
Table 2.3: Average number of hours per week in childcare

	Mean	Standard deviation
Hours in formal childcare	26.79	13.03
Hours in informal childcare	16.78	15.15

To conclude the descriptive analysis, Figure 2.2 documents the association between the average number of reported difficulties across all years surveyed and hours spent in formal childcare. As expected, the correlation is negative. Spending up to 8 hours per week in childcare is associated with an average of 6.8 reported difficulties, spending 40 hours a week in childcare is associated with an average of 5.7 reported difficulties.

Appendix Figures A.1 - A.5 show the correlation by each surveyed year. Reported difficulties reduce over time and the correlation gets weaker.

Figure 2.2: Descriptive relationship between hours in childcare and non-cognitive skills between age 3-14



Notes: Correlation between hours per week in formal childcare by age 3 and average reported difficulties from age 3-14. Reported difficulties are collapsed on the integer of hours in childcare. The size of the marker indicates the relative number of observations in the hours cell. The fitted line is taken from a simple linear regression of reported difficulties on weekly hours in childcare.

2.5.2 Regression Results

Table 2.4 contains the OLS estimates for non-cognitive skills from age 3-14 from the production function specified in equation (1). The OLS results are small in magnitude and not statistically significant. As discussed in Section 2.4, these estimates are likely biased. Previous childcare research (see Section 2.2) has found their unbiased estimates larger than OLS, suggesting a bias towards zero.[34]

Table 2.4: OLS estimation

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.001	0.001	0.001	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	6387	6567	5926	5616	5117
R^2	0.18	0.18	0.17	0.15	0.13

Notes: The dependent variable is the total number of reported difficulties at each age surveyed. The dependent variable is standardised with a mean of zero and a standard deviation of one. Hours is the number of hours per week in formal childcare by age 3. Controls include gender, ethnicity, month of birth, SEN, birth weight, development at 9 months old, mother's age at birth, whether the child has older siblings, mother's education, mother's economic status, household income, mother's health, presence of father and region of residence. All controls are measured at age 9 months. Standard errors are in parentheses. We assume iid errors.[a] *The change in sample size is due to non-response and attrition. The impacts of the change in the sample size are examined in the robustness checks.*

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] We also estimated all models with non-iid errors and results remained similar.

Turning to the IV results, the instruments perform well.[35] The first-stage F statistics (Cragg-Donald Wald F statistic) range from 31.1-23.5. Tests for both under and over identification are satisfied in each model. The weak instrument (Anderson-Rubin) robustness test of the joint significance of the instruments in the reduced form model is satisfied for non-cognitive skills at ages 3-7 but not age 11 and 14.

The first stage estimates (see Appendix Table A.4) suggest that a 10 percentage points

[34] The OLS estimated coefficients turn insignificant after controlling for mother's education.

[35] See robustness checks for the further discussion on the strength of the instrumental variables.

increase in the probability of the mother working shift work, reduces hours in formal childcare by 51.6 minutes per week. For a 10-percentage point increase in the probability of having uncertain hours, hours in formal care increase by 44.4 minutes per week.

The second-stage IV results, shown in Table 2.5[36], are larger in magnitude than the OLS estimates. The IV results indicate that on average, increasing hours in formal childcare at age 3 by 1, reduces reported difficulties at age 3 by 5.2% of a standard deviation. This reduces to 2.4% of a standard deviation by age 5, followed by 4.4% at age 7[37] and 2.3% and 2.1% at age 11 and 14 respectively.[38] Another way to interpret these findings is that increasing childcare by age 3 by half a day (4 hours) per week suggests that reported difficulties would reduce by slightly less than 1. Large findings are not uncommon within the literature. Felfe et al. (2015), Gupta & Simonsen (2016)and Berger et al. (2021) all estimate impacts of attending childcare from 15-34% of a standard deviation.

[36]Full model second stage results are shown in Table A.5 in the appendix.

[37]The reduction in the impact at age 5 followed by an increase in the estimates impact at age 7 is consistent with some of the literature. Peter et al. (2016), who also use data from the Millennium Cohort Study, report similar findings. Later day care entry significantly increases children's socio-emotional problems (SDQ) in the medium run (at age 7) but not in the short run (at age 5). As children enter school at age 5, there is a suggestion that such a big event in a child's life makes it difficult for the parent to correctly rank their difficulties.

[38]Due to the weak instrument robustness checks not being satisfied for age 11 and 14, the results here should be taken with caution. Further weak instrument testing is carried out in the robustness checks.

Table 2.5: Over-identified IV model

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.052***	-0.024*	-0.044***	-0.023	-0.021
	(0.014)	(0.013)	(0.012)	(0.014)	(0.014)
First Stage F	31.1	31.5	29.9	25.2	25.2
A-R Wald (P Value)	0.00	0.03	0.01	0.24	0.26
Hansan J Stat (P Value)	0.32	0.08	0.38	0.86	0.62
N	6387	6568	5926	5616	5117

Notes: The dependent variable is the total number of reported difficulties at each age surveyed. The dependent variable is standardised with a mean of zero and a standard deviation of one. Hours is the number of hours in formal childcare by age 3. Cragg-Donald Wald F statistic along the p-values associated with the Hansan J overidentification test and the weak instrument (Anderson-Rubin) robustness test are shown. Standard errors are in parentheses. Controls include gender, ethnicity, month of birth, SEN, birth weight, development at 9 months old, mother's age at birth, whether the child has older siblings, mother's education, mother's economic status, household income, mother's health, presence of father and region of residence. All controls are measured at age 9 months. The change in sample size is due to non-response and attrition. The impacts of the change in the sample size are examined in the robustness checks. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The difference between the OLS and IV results may reflect that children who are exogenously induced into hours in childcare based on their mother's work patterns are more heavily impacted by formal childcare than children whose parents select them into hours in childcare regardless of their work patterns.

The overall pattern of results suggests that any bias induced by the endogeneity of hours in formal childcare likely results in underestimation of the positive effect of hours in formal childcare on non-cognitive skills. As such, the IV estimation can be viewed as confirming the direction of the OLS results.[39]

[39] See 2.5.5 for discussion on the coefficient of proportionality.

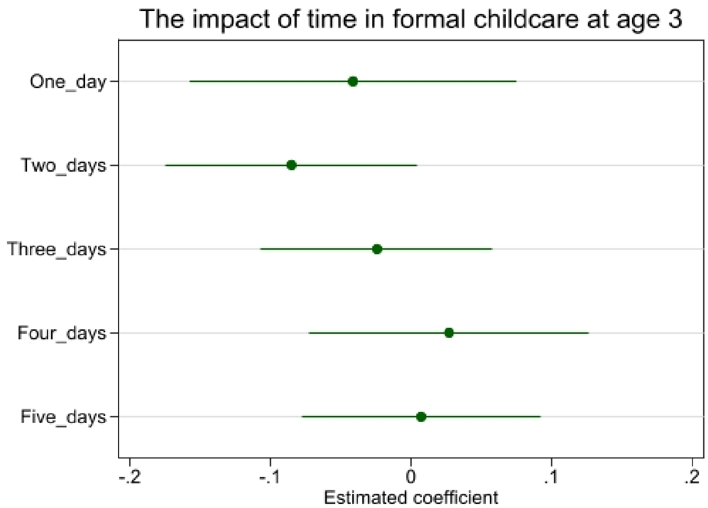
2.5.3 Functional Form

The simplifying assumption made in the baseline model is that there are neither diminishing nor increasing benefits to hours in childcare.

The next stage of the analysis is to explore the extent to which hours in childcare have non-linear effects on non-cognitive development. We may expect the positive impacts to increase over time as children spend more time in a structured setting with experienced carers and children of a similar age. On the other hand, if reduced time with parents has a negative impact on children, then we may also expect the non-linear relationship to work in the opposite direction. To investigate the presence of a non-linear effect, we use the measure of hours to generate dummy variables for days in formal childcare. We estimate an OLS specification with the dummy variables.[40]

[40] We use the OLS specification to examine the non-linear impact as we have been unsuccessful in finding an instrument for quadratic of hours in childcare at age 3 that satisfies weak instrument testing.

Figure 2.3: The non-linear impact of time in childcare at age 3 (OLS estimation)



Notes: These coefficients are estimated using an OLS regression. The dependent variable is the total number of reported difficulties at age 3. The dependent variable is standardised with a mean of zero and a standard deviation of one. The number of hours spent in childcare is used to calculate dummy variables for days in childcare. The reference category is no childcare. Controls include gender, ethnicity, month of birth, SEN, birth weight, development at 9 months old, household members, mother's age at birth, mother's education, household income, whether the child has older siblings, if the mother is a single parent and mother's health. All controls are measured at age 9 months.

Figure 2.3 suggests that the optimal amount of time to spend in formal childcare is two days. Beyond two days the impact decreases. We are cautious in the interpretation of these findings as these coefficients are estimated using an OLS estimation, which we expect to be biased downwards, we take these estimates as a lower bound.

2.5.4 Heterogeneity

On average, increasing the amount of time spent in formal childcare has a positive impact on non-cognitive development both in the short and medium term. The next stage of this analysis explores the extent to which the effect of hours in formal childcare differs across observable characteristics. Understanding heterogeneous impacts is important for policy development as it can identify groups which may benefit the most from the intervention.

The Strength and Difficulties questionnaire asks about 25 attributes, some positive and others negative. These 25 items are divided between 5 scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. For the baseline estimates we use total difficulties as the dependent variable. We start examining the heterogeneous impact by estimating the impact of hours in childcare across the 5 scales.[41]

Table 2.6: Impact of hours in formal care on type of non-cognitive skill estimated by an over-identified IV model

	(Emotion)	(Conduct)	(Hyper)	(Peer)	(Pro-social)
Impact at age 3	-0.022*	-0.041***	-0.034***	-0.086***	0.007
	(0.012)	(0.013)	(0.012)	(0.013)	(0.009)
Impact at age 5	-0.012	-0.028**	-0.012	-0.019	0.006
	(0.013)	(0.013)	(0.012)	(0.012)	(0.009)
Impact at age 7	-0.041***	-0.039***	-0.016	-0.039***	-0.002
	(0.013)	(0.014)	(0.012)	(0.013)	(0.009)
Impact at age 11	-0.019	-0.022	-0.008	-0.024*	-0.016
	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)
Impact at age 14	-0.015	-0.029**	-0.018	-0.001	0.001
	(0.014)	(0.015)	(0.014)	(0.014)	(0.011)

Notes: The dependent variable is the total number of reported difficulties in each sub-section at each age. The dependent variables are standardised with a mean of zero and a standard deviation of one. Hours is the number of hours in formal childcare by age 3. Standard errors in parentheses. Controls include gender, ethnicity, month of birth, SEN, birth weight, development at 9 months old, mother's age at birth, whether the child has older siblings, mother's education, mother's economic status, household income, mother's health, presence of father and region of residence. All controls are measured at age 9 months. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We find no significant impact on pro-social behaviour. Pro-social behaviour is not included in total difficulties and is a measure of positive non-cognitive skills. We therefore argue that time in childcare at age 3, improves (i.e. reduces) negative non-cognitive skills but

[41] First stage F statistics range from 31.1-25.2 and over-identification checks are valid for all models.

has no effect on positive non-cognitive skills. We find significant improvements in all other measures of strength and difficulties for age 3 and all, expect hyper, for age 7. The impact on emotional skills, conduct and peer relationships could be argued to be the most persistent over time. Difficulties related to emotional skills are worrying, being nervous and scared. For conduct, difficulties include getting into fights, getting angry and not doing as they have been told. In relation to peer relationship, difficulties include being solitary and getting on better with adults. Spending more time in childcare at a young age could be argued to increase independence, get children used to being out of their comfort zone, increasing interaction with other children and being in a more structured environment where people are trained to help children understand their emotions. These findings are partially in line with Morris et al. (2021A) who also find positive impacts of formal childcare usage on emotional and peer problems. However, they also find positive impacts on pro-social behaviour and negative impacts on conduct problems. These differences could be explained by the research setting, Morris et al. (2021A) focus on disadvantaged children whilst this research takes a whole population view.

We also conduct a series of subgroup analyses based on individual and family characteristics; the results are shown below.[42]

We start by examining the impact across gender.

[42]We estimate all subgroup analysis using the over-identified IV model. First stage F statistics remain above 10 for all subgroups and overidentification tests are satisfied.

Table 2.7: The impact of hours in childcare by age 3 on total difficulties, by gender. (Estimated by an over-identified IV model)

	(Female)	(Male)
Impact at age 3	-0.041**	-0.067***
	(0.017)	(0.022)
Impact at age 5	-0.027*	-0.027
	(0.016)	(0.019)
Impact at age 7	-0.041**	-0.048**
	(0.018)	(0.020)
Impact at age 11	-0.012	-0.039*
	(0.018)	(0.023)
Impact at age 14	0.001	-0.059**
	(0.017)	(0.026)

Notes: The dependent variable is the total number of reported difficulties at each age surveyed. The dependent variable is standardised with a mean of zero and a standard deviation of one. Hours is the number of hours in formal childcare by age 3. Standard errors are in parentheses. Controls include gender, ethnicity, month of birth, SEN, birth weight, development at 9 months old, mother's age at birth, whether the child has older siblings, mother's education, mother's economic status, household income, mother's health, presence of father and region of residence. All controls are measured at age 9 months. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

With respect to gender, we find time spent in formal childcare improves non-cognitive skills for both males and females. We estimate a larger impact for males which is persistent over time.

Blanden et al. (2022), who also use English data, find no significant impact for females. Studies looking at other European countries for example, Felfe et al. (2015), find larger impacts for females.

We also examine whether there is a difference in impact depending on if the child has older siblings. We could assume that children who are first born may be most impacted by time in childcare as their parents have less experience. On the other hand, children with older siblings may be impacted more by increasing time in childcare as they have to share their parents' time with their siblings. Table 2.8 provides some suggestions that it is children who are not first born which seem to benefit the most from spending more time in childcare, this effect is also persistent over time.

Table 2.8: The impact of hours in childcare by age 3 on total difficulties, by first born status. (Estimated by an over-identified IV model)

	(First Born)	(Older Siblings)
Impact at age 3	-0.039***	-0.055**
	(0.014)	(0.024)
Impact at age 5	-0.023*	-0.019
	(0.013)	(0.023)
Impact at age 7	-0.031**	-0.046*
	(0.014)	(0.024)
Impact at age 11	-0.014	-0.038
	(0.015)	(0.026)
Impact at age 14	-0.004	-0.041*
	(0.014)	(0.027)

Notes: See Table 2.7. At age 3, 43% of the sample were first born.

There is some degree of consensus in the literature that childcare is more beneficial for disadvantaged children than for their more advantaged counterparts. To examine this, we carry out subsample analyses using mother's education[43] and equivaled household income.[44] Table 2.9 present the results for mother's education whilst Table 2.10 shows the estimated impact across the household income distribution.

[43] Mother's education is a dummy variable equal to one if she has a degree or equivalent

[44] To examine the impact of across the income distribution we generate three dummy variables relating to the bottom 20%, top 20% and middle 60% of the distribution.

Table 2.9: The impact of hours in childcare by age 3 on total difficulties, by mother's education. (Estimated by an over-identified IV model)

	(Degree or higher)	(No degree)
Impact at age 3	-0.034**	-0.064***
	(0.015)	(0.019)
Impact at age 5	-0.033**	-0.025
	(0.016)	(0.017)
Impact at age 7	-0.042**	-0.044**
	(0.017)	(0.018)
Impact at age 11	-0.016	-0.019
	(0.016)	(0.019)
Impact at age 14	-0.017	-0.013
	(0.015)	(0.019)

Notes: See Table 2.7. At age 3, 22% of the sample had mothers with a degree.

Whilst we find that the initial impact of hours in childcare is larger for children with mothers who have lower levels of education, this impact declines over time and becomes insignificant. By age 5, there is no significant difference in the impact. We could therefore argue that children who have lower educated mothers do benefit greatly from time spent in childcare however, children with higher educated mothers are better at maintaining the skills learnt.[45]

Table 2.10: The impact of hours in childcare by age 3 on total difficulties, by household income. (Estimated by an over-identified IV model)

	(Top 20%)	(Middle 60%)	(Bottom 20%)
Impact at age 3	-0.026**	-0.067 ***	-0.024*
	(0.012)	(0.022)	(0.129)
Impact at age 5	-0.030**	-0.042**	-0.090
	(0.012)	(0.019)	(0.080)
Impact at age 7	-0.018	-0.056***	-0.193**
	(0.012)	(0.021)	(0.091)
Impact at age 11	-0.003	-0.039*	-0.090
	(0.012)	(0.021)	(0.189)
Impact at age 14	-0.003	-0.029	0.003
	(0.012)	(0.020)	(0.160)

Notes: See Table 2.7

[45] Table A.6 in the appendix presents a sub-sample analysis for mothers with qualifications higher than the school level. The analysis reveals that the impact of time spent in formal childcare is larger for children whose mothers have only school-level qualifications, and this impact persists over time.

Regarding household income, we find larger impacts on children in households in the middle of the income distribution compared to individuals in the top 20% of the distribution. This provides suggestive evidence that individual children from less advantaged backgrounds benefit more from time in childcare at a young age.

We interpret the impact for those in the bottom 20% of the distribution with caution as children from households with higher income are more likely to enrol in childcare for more hours and therefore, the sample of children in the bottom 20% of the distribution which has hours in childcare larger than zero is small. The instruments are weak for the sub-sample analysis on the bottom 20% so we focus only on the top and middle of the income distribution for our interpretation. Since the bias will be in the direction of the OLS we can interpret these estimates as lower bounds.

2.5.5 Robustness Check

Weak instrument robustness check

An instrumental variable is categorised as weak if the correlation with the endogenous variable, conditional on any controls, is close to zero. When this correlation is low, the usual approximations for the distribution of IV estimators become unreliable, leading to biased IV estimates and confidence intervals that are less likely to contain the true parameter value.

Recognising this issue has led to extensive research on econometric methods for models with weak instruments. Staiger & Stock (1994) proposed a higher standard of instrument relevance in the first stage of 2SLS. They show that to be confident the estimator behaves as intended, instrument significance should be higher than the conventional 5% level in the first stage. Their findings indicate that if the first-stage F statistic exceeds 10 in the one endogenous variable, one instrument case, then 2SLS two-tailed t-tests will reject a true null

$H_0 : \beta = 0$ at a rate close to the correct 5% rate.

The advice to have an F statistics greater than 10 has been widely adopted in practice despite many theoretical analysts arguing for higher levels of acceptance. Andrews et al. (2019) compared the F statistics of 17 papers published in the American Economic Review from 2014 to 2018 that use instrumental variables. They show that the majority of studies present F statistics between 10-15 with some even below 10.

Most recently, Keane & Neal (2021) has demonstrated that 2SLS standard errors are often small in samples where the estimates are significantly affected by the OLS bias. This results in an inflated power of the t-test to detect false positive effects when the OLS bias is positive. This issue remains even when the first-stage F statistics is very high.

They further show that the weak instrument robustness test proposed by Anderson & Rubin (1949) improves the power asymmetry problem and is therefore more reliable than the t-test for assessing the significance of the instrumenting variable's coefficient. The AR test is recommended for use instead of the t-test, even when instruments are strong.

The first stage F statistics in this research range between 31-25, higher than many studies discussed in Andrews et al. (2019). Using the Stock-Yogo critical values, we can reject the hypothesis that the relative (vs. OLS) asymptotic bias could be greater than 10%.

Table 2.11 presents the weak instrument robust 95% confidence intervals. The weak instrument robust testing supports the findings in the baseline estimates. The weak instrument (Anderson-Rubin) robustness test of the joint significance of the instruments in the reduced-form model is satisfied for non-cognitive skills at ages 3-7 but not age 11 and 14. When looking at the robust confidence intervals, we find significant impacts for children aged 3 to 7 however the impact declines over time and slips into insignificance at ages 11 and 14. We therefore argue that hours in childcare has a significant impact on initial non-cognitive skills which persists in the medium term. However, the findings for the long run are less clear.

Table 2.11: Weak instrument robust confidence intervals

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	[-0.089,-0.025]	[-0.049,-0.004]	[-0.082,-0.016]	[-0.064, 0.010]	[-0.061, 0.011]

Notes: Anderson and Rubin 95% confidence intervals for IV coefficients presented in Table 2.5. The dependent variable is standardised with a mean of zero and a standard deviation of one.

Just identified model

Adding more instruments increases asymptotic efficiency of the 2SLS estimator. The finite sample bias in 2SLS can however, get much worse by adding too many instruments (Bound et al. 1995).

Since Angrist & Kolesár (2024) showed that just-identified IV estimated models perform well even with weak instruments unless the degree of endogeneity is very high, as a robustness check we estimate a just-identified model using each of our instruments separately.

Table 2.12 presents the IV estimates of a just identified model using the probability of mother's working shift work as an instrument for hours in formal childcare. We find very similar estimates to our baseline model, though slightly smaller in magnitude. Table 2.13 shows the IV estimates of a just identified model using the probability of the mother having uncertain hours as an instrument for hours in formal childcare. The estimated coefficients are again similar to our baseline estimates though become insignificant at age 7. We should take note that the first stage F statistic is a lot smaller in the latter case.

Table 2.12: Just identified model using probability of shift working at an instrument

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.042**	-0.008	-0.036**	-0.021	-0.017
	(0.016)	(0.015)	(0.016)	(0.017)	(0.016)
First Stage F	40.1	41.1	39.4	34.4	36.9
N	6387	6568	5926	5616	5117

Notes: See Table 2.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.13: Just identified model using probability of working uncertain hours

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.055*	-0.059*	-0.042	-0.012	-0.026
	(0.031)	(0.034)	(0.029)	(0.027)	(0.036)
First Stage F	8.9	7.9	8.5	8.6	5.1
N	6387	6568	5926	5616	5117

*Notes: See Table 2.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Sample Selection

The Millennium Cohort Study, like many other longitudinal surveys, experiences issues with non-response and attrition. The concern is that cases that are lost, may differ in characteristics from those that remain. As a result, inferences drawn from the observed sample may not reflect the population as a whole. If parents who are more likely to respond to the survey are also more likely to enrol their child in formal childcare for more hours, then this will create a bias.

The Millennium Cohort Study has faced large attrition. During wave 1, at age 9 months, they surveyed 18,818 individuals. This dropped to 11,872 individuals by age 14. Many individuals also have gaps in their responses, some missed waves due to illness for example and many did not respond to certain questions.

Plewis (2007) analyses factors influencing non-response between the first and second wave of the Millennium Cohort Study. They provide evidence to show that young mothers as well as breast-feeding mothers, respondents from minority groups, notably Black and 'other' minority ethnic groups, respondents with fewer educational qualifications, poorer families, living in rented accommodation and not in a house are more likely to leave the sample.

This research accounts for many of these factors, and Plewis (2007) contends that although the individuals lost from the sample differed from those who remained, the differences were not significantly substantial, as far as could be determined.

To examine whether changes in the sample size are driving the estimated results, we restrict the sample to individuals who have responded to every wave.

Table 2.14 shows that whilst the estimated coefficients are smaller than the baseline, the pattern remains the same. Time in formal childcare has positive impacts on non-cognitive skills, the impact reduced though remains significant at age 7.[46]

Table 2.14: Estimates from over-identified IV model on sample of individuals who responded to all waves

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.047***	-0.018	-0.036**	-0.018	-0.020
	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)
N	4148	4148	4148	4148	4148

Notes: See Table 2.7 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Coefficient of proportionality

The IV estimates presented in Table 2.5 are larger than the OLS estimates presented in Table 2.4. As we expect that any bias induced by the endogeneity of hours in formal childcare likely results in underestimation of the positive effect of hours in formal childcare on non-cognitive skills, this is what we would have expected.

Taking the estimated coefficient on non-cognitive skills at age 3 as an example, here we estimate IV coefficients nearly 6 times larger than the OLS model. Large differences between the size of OLS and IV coefficients can be a potential indicator of instrument validly issues.

Ciacci (2021) argues that solely comparing the relative size of OLS and IV estimates is insufficient to determine whether the IV estimate reliably reflects the true effect. They suggest computing the coefficient of proportionality using Oster (2019) bounds to compare these two

[46] The instruments remain strong with a first stage F statistic of 22.1.

estimates. The coefficient of proportionality tells us how strong selection on unobservables must be compared to selection on observables, to estimate the IV coefficient with the OLS model. Ciacci (2021) suggests that low values of the coefficient of proportionality provide supportive evidence that IV estimates are not too large with respect to OLS.

Table 2.15 presents the coefficients estimated by Oster (2019) methodology setting a negative sign of the coefficient of proportionality γ and Rmax = 1.

Table 2.15: Coefficient of proportionality

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
γ	-1.91	-0.72	-3.55	-0.73	-0.49

Notes: The output of the OLS regression is displayed in Table 2.4. The output of the IV regression is displayed in Table 2.5.

At first sight the difference between OLS and IV estimates might appear considerable, however, taking into account Table 2.15, Oster (2019) methodology finds that a negative coefficient of proportionality with size between 3.55 and 0.49 is enough to identify a set that includes the IV estimates. As long as selection on unobservables is at the largest 3.6% larger than selection on observables it is enough for the true treatment effect to have the size of the IV estimates.

Sure Start

Starting in 2000, the Department for Education allocated funds for childcare places for three-year-olds in 65 Local Education Authorities (LEAs). By 2001, this initiative was expanded nationwide, with the goal of achieving universal coverage for all three-year-olds by 2004. This means that during the period in focus, not all children aged three and younger were eligible for any government childcare funding. However, Sure Start, introduced in 1999, was the first major government initiative to provide comprehensive support to families with children under five in England. The policy created a network of local 'one-stop shops' that offered a variety of services aimed at improving the development and life chances of young children. These services included health services, parenting support, early learning and childcare, and assistance with parental employment. Up until 2004, Sure Start focused on drop-in sessions and activities for parents, carers, and children. Sure Start offered childcare after 2004.

Within the MCS, parents were asked in Wave 1, when their child was 9 months old, if they had used Sure Start services. Since this was a time when the programme was still in its early stages, the majority of respondents had not used it. Due to the small sample size of those who had, we are unable to determine any differences in impact between the two groups. However, by focusing our analysis on a sample of individuals who had not used these services in 2000, we can demonstrate that our results are not influenced by those who had.

Table 2.16 shows the estimated impact of hours in childcare on non-cognitive skills at age 3 who had not used Sure Start when the child was 9 months old. Whilst the estimated coefficient is slightly smaller than the baseline model, we still find that increasing hours in childcare reduces total difficulties and improves non-cognitive skills.

Table 2.16: Over identified model on a sample of individuals who did not use Sure Start

	(Non-cog 3)
Hours	-0.035 ***
	(0.012)
First Stage F	31.7
N	4222

*Notes: See Table 2.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Omitted variables

All specifications control for individual and parental characteristics when the individual was 9 months old. Despite arguably controlling for a good amount of selection into childcare, there is still potential for selection on unobservables. One of the main potential unobservable factors is parental preferences for childcare. Whilst many studies use household income and education measurements as a proxy, Bernal & Keane (2010) argue that the ability to provide does not always lead to provision, and that differences are potentially driven by preferences.

When the child is 9 months old, the mother is asked whether she believes working before the child enters school is detrimental to the child. The responses are recorded in a categorical variable ranging from strongly agree to strongly disagree. We use this categorical variable as a proxy for preferences.

Table 2.17 presents the IV estimates with this additional control.[47] The results are very robust to this controlling for preferences, with only a very small decrease in the magnitude of the estimates.

Table 2.17: Over-identified IV model including measure of mother's preferences for childcare

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.051***	-0.024*	-0.034***	-0.023	-0.021
	(0.013)	(0.012)	(0.014)	(0.014)	(0.014)
N	6387	6568	5926	5616	5117

Notes: See Table 2.7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Another potential omitted variable is distance from grandparents. Being cared for by the maternal grandmother is the most common type of informal childcare therefore the distance from grandparents is likely to be a large determinant of the amount of hours children spend in formal childcare. Studies including Hansen & Hawkes (2009), also show grandparent care to have an influence on child development.

Table 2.18 presents the IV estimates with this additional control. The results are again very robust to the controlling of distance lived from grandparents.

[47] First stage F statistics remain between 27-21.

Table 2.18: Over-identified IV model including measure of distance to grandparents

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.053***	-0.027**	-0.043***	-0.030*	-0.027*
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
N	5482	5656	5099	4819	4362

Notes: See Table 2.7 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Mother's job

The validity of our instruments assumes that whilst the probability of working shifts and having uncertain hours are correlated with the number of hours the mother enrols her child in formal childcare for, they are uncorrelated with their non-cognitive developments.

To try to reduce the concern that mothers select in to jobs based on their child's non-cognitive skills, we use the characteristics of the job mothers had during pregnancy as the instrument due to it being pre-determined.

As a robustness check, we examine the strength of the instrumental variables further by utilising some of the extra information in the survey.

Firstly, whilst we control for whether the mother was employed or not when the child was 9 months old, since 24% of mothers return to work part-time after having children, we can additionally control for the number of hours worked. The results reported in Table 2.19 are very similar to the baseline results reported in Table 2.5.

Table 2.19: Over-identified IV model including hours worked

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.054***	-0.024*	-0.047***	-0.027*	-0.023
	(0.015)	(0.013)	(0.015)	(0.015)	(0.015)
N	6387	6568	5926	5616	5117

Notes: See Table 2.7 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Secondly, parents are asked why they left their last job as well as what different characteristics their new job has. We drop observations where mothers report leaving a job due to not having flexible hours. We also drop observations where mothers report having a change in hours in their new job. Despite the change in sample, the results reported in Table 2.20 are very similar to the baseline results reported in Table 2.5.[48]

Table 2.20: Over-identified IV model dropping mother's who change job due to hours worked

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.061***	-0.025*	-0.044***	-0.030*	-0.026*
	(0.016)	(0.014)	(0.016)	(0.016)	(0.016)
N	5634	5804	5221	4947	4521

Notes: See Table 2.7 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, we drop observation were the mother changes SOC code between pregnancy and the child being three years old. Table 2.21 presents results which are consistent with the baseline findings.

Table 2.21: Over-identified IV model dropping mother's who change SOC

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.063***	-0.031**	-0.054***	-0.023	-0.025
	(0.017)	(0.016)	(0.018)	(0.018)	(0.018)
N	4910	5075	4565	4309	3939

Notes: See Table 2.7 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6 Conclusion

This research provides the first attempt at identifying the causal relationship between hours in childcare prior to the age of 3 and non-cognitive skills reported at ages 3-14. We aim to shed light on the causal relationship by adopting an instrumental variables strategy that leverages exogenous variation in both the probability that the mother works shift work and

[48]The first stage F stats remain at 26-23.

has uncertain working hours. Our IV results suggest that the relationship between hours in childcare and non-cognitive skills is likely causal and that the more naive OLS estimates likely underestimate the causal effect.

Using the Millennium Cohort Study which focuses on children born in the UK in the year 2000, we find that increasing hours in formal childcare by age 3 by 1, reduces reported difficulties at age 3 by 5.2% of a standard deviation. This reduces to 2.4% of a standard deviation by age 5, followed by 4.4% at age 7. We also find smaller impacts of 2.3% and 2.1% at age 11 and 14 respectively. These findings show that time spent in childcare by age 3 has initial impacts on non-cognitive skills which persist into the medium run. In respect to the long run, we are cautious in our interpretation as the impacts slip into insignificance when performing weak instrument robust testing. The fact that these effects do persist underlines the importance of early childhood education. We also find descriptive evidence of a non-linear relationship where the impact increases for up to two days in formal childcare and then decreases. We hypothesise that the result is potentially related to the mix of exposures children are getting when they spend time in childcare and being cared for by their parents.

Our findings, especially for the medium run impact, deliver very robust results across different samples of individuals and measures of non-cognitive skills. The robustness checks consistently suggest that the existence of weak instruments or omitted variable bias is rather unlikely.

How do our findings fit within the literature? We estimate positive effects of increased time in childcare on non-cognitive skills whereas Gupta & Simonsen (2010) and Berger et al. (2021) who provide descriptive evidence on the impact of time spent in childcare, both estimate negative effects.[49] Firstly, our estimates represent LATEs rather than average treatment effects. As such, they are less readily comparable to estimates from prior work. Secondly,

[49] There is a larger literature which examines the impact of attending childcare which also find mixed evidence.

the literature on the impact of hours spent in childcare is very small and for the most part descriptive, making it impossible to draw any general conclusions.

This research has useful policy implications. Firstly, consistent with the broader literature on childcare enrolment, we estimate positive effects of childcare. There is a current interest in the optimal number of hours in childcare in England due to the upcoming changes in childcare subsidies. The subsided hours in childcare for working parents are increasing from 30 hours for 3- and 4-year-olds to cover all under 5-year-olds by 2025. Whilst this policy was driven by the impact on female labour supply, our findings provide evidence to show that the policy change also has benefits for the child.

Secondly, results from our sub-group analysis suggest that there is heterogeneity in the magnitude and persistence of impacts across population subgroups. Most notably, less advantaged children—particularly those with low-educated mothers and those in lower-income households appear to initially benefit most from hours spent in childcare. These findings are consistent with a large proportion of the literature which focuses on childcare for disadvantaged children. These findings support the argument that facilitating access to more time in childcare for disadvantaged children may hold potential for decreasing early socioeconomic disparities in child development. Though we also find that these initial impacts do not persist, these findings require further investigation to understand whether disadvantaged children need more assistance to maintain skills learnt in early childcare.

These results should be interpreted in the context of a few limitations. Firstly, they are English specific and therefore should be generalised with caution. Secondly, children spending fewer hours in childcare could be in many other childcare settings for varying lengths of time and with varying quality.[50] [51] Finally, our IV analyses rely on the assumption that mothers do not select into occupations based on their child's non-cognitive skills. If this assumption

[50]Morris et al. (2018) show that the benefits of high quality childcare are larger when compared with parental care than other types of formal childcare.

[51]Due to sample size, we cannot explore this further.

is incorrect, our IV estimates will be biased.

In future research, it would be worthwhile to explore the impact across different care settings, extend the analyses to cognitive measures and investigate some of the mechanisms involved in the impact.

This research has contributed to the literature by investigating the causal relationship between hours in childcare and non-cognitive skills. We estimate positive findings which we argue suggests that an expansion of subsided hours in childcare will be beneficial for all, especially children from disadvantaged backgrounds, which may have potential to contribute to decreasing early gaps in child development.

Appendix A

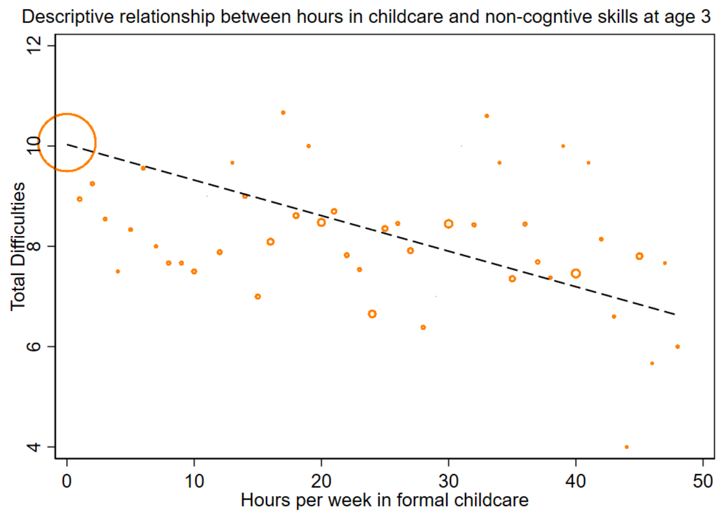
Appendix

Table A.1: Average number of reported difficulties

	Parent only care	Grandparent care	Formal childcare	Other
Total difficulties age 3	10.35 (5.51)	10.16 (5.36)	8.01 (4.30)	9.34 (4.80)

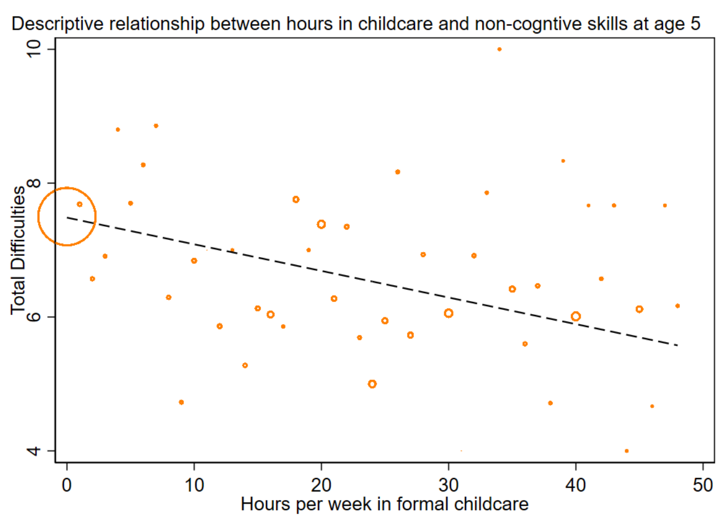
Children in formal childcare by age 3 have significantly lower total difficulties than all other types of childcare. Other types of childcare include friends, nannies and au pairs.

Figure A.1: Descriptive relationship between hours in childcare and non-cognitive skills at age 3



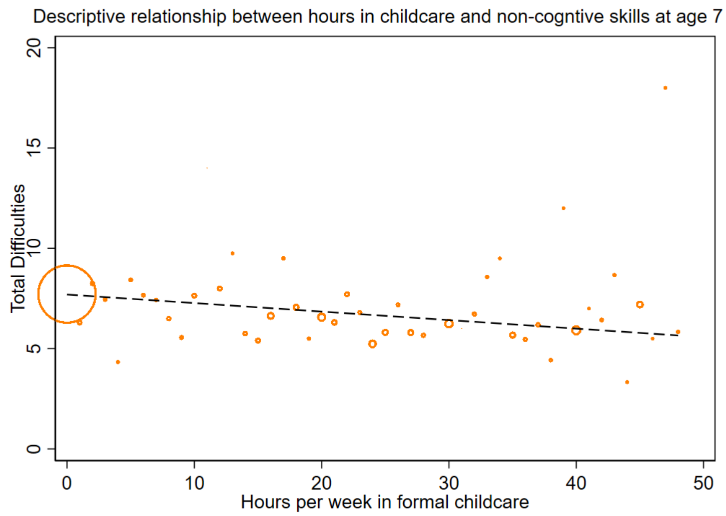
Notes: Correlation between hours per week in formal childcare by age 3 and average reported difficulties at age 3. Reported difficulties are collapsed on the integer of hours in childcare. The size of the marker indicates the relative number of observations in the hours cell. The fitted line is taken from a simple linear regression of reported difficulties on weekly hours in childcare.

Figure A.2: Descriptive relationship between hours in childcare and non-cognitive skills at age 5



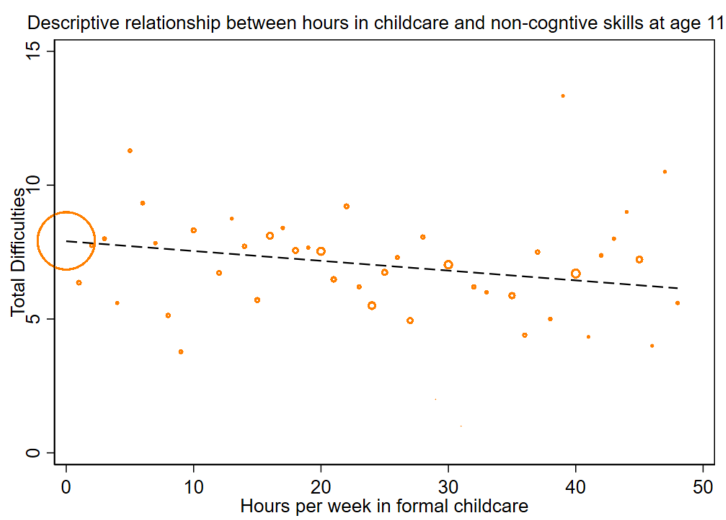
Notes: Correlation between hours per week in formal childcare by age 3 and average reported difficulties at age 5. Reported difficulties are collapsed on the integer of hours in childcare. The size of the marker indicates the relative number of observations in the hours cell. The fitted line is taken from a simple linear regression of reported difficulties on weekly hours in childcare.

Figure A.3: Descriptive relationship between hours in childcare and non-cognitive skills at age 7



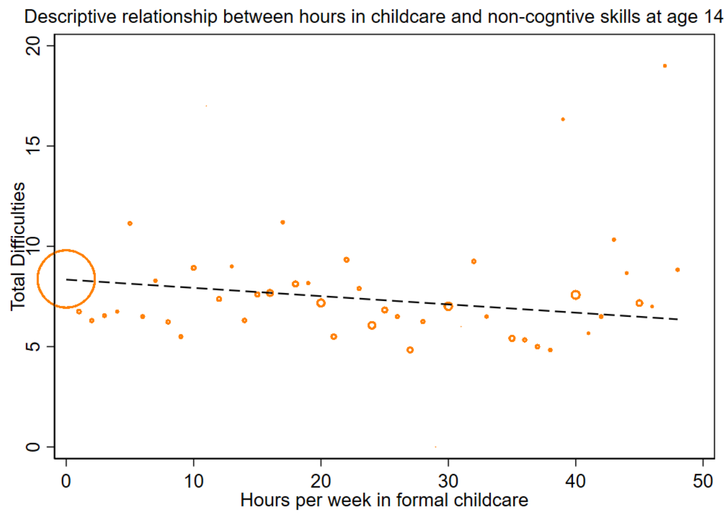
Notes: Correlation between hours per week in formal childcare by age 3 and average reported difficulties at age 7. Reported difficulties are collapsed on the integer of hours in childcare. The size of the marker indicates the relative number of observations in the hours cell. The fitted line is taken from a simple linear regression of reported difficulties on weekly hours in childcare.

Figure A.4: Descriptive relationship between hours in childcare and non-cognitive skills at age 11



Notes: Correlation between hours per week in formal childcare by age 3 and average reported difficulties at age 11. Reported difficulties are collapsed on the integer of hours in childcare. The size of the marker indicates the relative number of observations in the hours cell. The fitted line is taken from a simple linear regression of reported difficulties on weekly hours in childcare.

Figure A.5: Descriptive relationship between hours in childcare and non-cognitive skills at age 14



Notes: Correlation between hours per week in formal childcare by age 3 and average reported difficulties at age 14. Reported difficulties are collapsed on the integer of hours in childcare. The size of the marker indicates the relative number of observations in the hours cell. The fitted line is taken from a simple linear regression of reported difficulties on weekly hours in childcare.

Table A.2: Over-identified IV model using one-digit SOC codes as instrument

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.068***	-0.053***	-0.070***	-0.059***	-0.061***
	(0.017)	(0.016)	(0.019)	(0.019)	(0.022)
First Stage F	23.4	23.7	20.5	17.2	14.2
N	6387	6568	5926	5616	5117

Notes: See Table 2.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Over-identified IV model using two-digit SOC codes as instrument

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.031**	-0.024*	-0.041***	-0.021	-0.021
	(0.013)	(0.013)	(0.014)	(0.015)	(0.016)
First Stage F	27.7	28.3	26.5	21.8	20.3
N	6387	6568	5926	5616	5117

Notes: See Table 2.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: The impact of hours in childcare by age 3 on non-cognitive skills at age 3 (First Stage for 2SLS estimation using probability of shift work and uncertain working hours as the instrumental variables)

	(Hours)
Shift work (IV)	-0.057***
	(0.008)
Uncertain hours (IV)	0.047***
	(0.010)
Female	-0.335
	(0.244)
Mixed ethnicity	-0.593
	(1.165)
Asian ethnicity	0.191
	(0.439)
Black ethnicity	2.295***
	(0.687)
Other ethnicity	-0.619
	(0.918)
Birth weight	0.096
	(0.219)
SEN	-0.424
	(0.426)
Siblings	-0.630***
	(0.138)
Household income	0.007***
	(0.001)
Mother L1 education	-0.646
	(0.562)
Mother L2 education	-0.336
	(0.443)
Mother L3 education	0.428
	(0.509)
Mother L4 education	0.076
	(0.579)
Mother L5 education	1.969***
	(0.534)
Mother L6 education	3.554***
	(0.799)
Mother other education	1.109*
	(0.568)
Main Employed	1.247
	(0.915)
Main Self employed	-2.475**
	(1.027)
Main out of work	-0.753
	(1.198)
Mother age	-0.106***
	(0.026)
Development 1	0.233
	(0.281)
Development 2	0.094
	(0.149)

Development 3	-0.358
	(0.456)
Yorkshire	0.405
	(0.633)
East Midlands	0.952
	(0.666)
West Midlands	1.019
	(0.640)
East of England	1.035
	(0.646)
London	0.015
	(0.628)
South East	0.612
	(0.608)
South West	0.337
	(0.677)
February	-0.171
	(0.601)
March	-0.399
	(0.572)
April	0.980
	(0.583)
May	0.020
	(0.574)
June	0.928
	(0.575)
July	0.540
	(0.584)
August	0.741
	(0.594)
September	-0.076
	(0.569)
October	-0.040
	(0.574)
November	0.312
	(0.577)
December	0.133
	(0.564)
Single Parent	2.838***
	(0.389)
Health good	-0.178
	(0.273)
Health fair	0.402
	(0.397)
Health poor	0.313
	(0.803)
N	6387

Notes: First stage regression for hours in childcare by age 3. All controls are measured at age 9 months. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Second stage over-identified IV model of the impact of hours in formal childcare by age 3 on non-cogntive skills at age 3 (Over-identified model)

	(Non-cognitive age 3)
Hours	-0.052***
	(0.014)
Female	-0.174***
	(0.027)
Mixed ethnicity	-0.066
	(0.125)
Asian ethnicity	0.302***
	(0.046)
Black ethnicity	0.071
	(0.078)
Other ethnicity	0.018
	(0.099)
Birth weight	-0.019
	(0.024)
SEN	0.322***
	(0.046)
Siblings	-0.055***
	(0.018)
Household income	-0.000
	(0.000)
Mother age	0.012***
	(0.003)
Mother L1 education	-0.155**
	(0.060)
Mother L2 education	-0.238***
	(0.047)
Mother L3 education	-0.249***
	(0.055)
Mother L4 education	-0.316***
	(0.061)
Mother L5 education	-0.284***
	(0.067)
Mother L6 education	-0.179*
	(0.104)
Mother other education	-0.134**
	(0.064)
Main Employed	0.095
	(0.099)
Main Self employed	-0.164
	(0.114))
Main out of work	0.087
	(0.129)
Development 1	0.098***
	(0.030)
Development 2	0.064***
	(0.016)

Development 3	0.161***
	(0.049)
North West	0.001
	(0.697)
Yorkshire	0.105
	(0.068))
East Midlands	0.005
	(0.073)
West Midlands	0.080
	(0.070)
East of England	-0.055
	(0.071)
London	-0.112*
	(0.067)
South East	-0.051
	(0.066)
South West	-0.044
	(0.073)
February	0.005
	(0.064)
March	0.017
	(0.061)
April	-0.014
	(0.064)
May	0.013
	(0.061)
June	0.023
	(0.061)
July	0.038
	(0.063)
August	0.039
	(0.064)
September	0.019
	(0.061)
October	0.035
	(0.061)
November	0.014
	(0.062)
December	0.037
	(0.060)
Single Parent	0.342***
	(0.056)
Health good	0.122***
	(0.029)
Health fair	0.297***
	(0.043)
Health poor	0.597***
	(0.086)
N	6387

Notes: The dependent variable is the total number of reported difficulties at age 3. All controls are measured at age 9 months. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: The impact of hours in childcare by age 3 on total difficulties, by mother's education. (Estimated by an over-identified IV model)

	(Higher than school level)	(School level)
Impact at age 3	-0.050	-0.100 ***
	(0.034)	(0.036)
Impact at age 5	-0.039	-0.051
	(0.031)	(0.033)
Impact at age 7	-0.057 *	-0.087 **
	(0.034)	(0.037)
Impact at age 11	-0.022	-0.036
	(0.032)	(0.035)
Impact at age 14	-0.003	-0.054
	(0.027)	(0.036)

Notes: See Table 2.7.

Table A.7: Over-identified IV model (paid for hours in childcare)

	(Non-cog 3)	(Non-cog 5)	(Non-cog 7)	(Non-cog 11)	(Non-cog 14)
Hours	-0.044***	-0.022	-0.038***	-0.009	-0.015
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
First Stage F	38.5	39.6	38.9	35.4	32.1
N	5062	5193	4645	4415	4007

Notes: The dependent variable is the total number of reported difficulties at each age surveyed. The dependent variable is standardised with a mean of zero and a standard deviation of one. Hours is the number of paid hours in formal childcare by age 3. Cragg-Donald Wald F statistic is shown. Standard errors are in parentheses. Controls include gender, ethnicity, month of birth, SEN, birth weight, development at 9 months old, mother's age at birth, whether the child has older siblings, mother's education, mother's economic status, household income, mother's health, presence of father and region of residence. All controls are measured at age 9 months. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Does one day make a difference? The short run impact of absences on individual outcomes in England

3.1 Introduction

Time spent in the classroom has been described as the single most powerful tool to make sure every child fulfils their potential (Department for Education 2022A). When pupils are absent from school, they not only miss learning but also peer and teacher interactions which are all seen to be important for pupil outcomes. The importance of attendance is well-documented in the correlation evidence between pupil absences and education outcomes (Department for Education 2016). Despite large policy focus, there is limited causal evidence, specifically outside the US, making effective policy design and targeting of intervention challenging. This research aims to shed light on the causal relationship between school absences and individual achievement for pupils in England. Using novel longitudinal data by matching individual absence and achievement data to survey responses, we show that even a small number of absences significantly reduce academic performance in Maths and English.

The persistently high absence rates across the world are being described as an empty desk epidemic. Whilst it was assumed that school absences would quickly fall back to their pre-Covid level, absence rates continue to be high four years later. In England, the absence rate for 2018/19 was 4.7%, which equates to 59.6 million days lost (Department for Education 2019). In 2022/23, absences rose by 60% to 7.5% of school days being missed (Eyles et al. 2023). We see similar evidence in America as more than 25% of pupils missed at least 10% of the 2021/22 school year in comparison to 15% prior to the pandemic (Dee 2024). The picture is also similar across Europe, for example in Belgium school absences increased by 28% post pandemic (Taylor 2023). We are now not only seeing persistently high levels of absences, but they are also not as socially stratified as prior to the pandemic.[1] Most pupils are absent more often, post-pandemic than pre. We are now potentially facing a worldwide education crisis in the post-pandemic era: the majority of pupils are not maximising the quantity of instructional time. With a rise in absences across the pupil population, it is now more important than ever to quantify the causal impact of absences on achievement.

This research aims to shed light on the causal relationship between pupil absences and educational achievement in the context of England. We construct a novel panel dataset by merging administrative data on pupil achievement and absences to a household survey which provides information on individual and family background. To the best of our knowledge, this is the first-time individual level absence data has been merged with individual survey responses. The use of survey responses over time means we have much richer background information on the pupils than previous studies in the literature. This means we rely less heavily on the assumption of time invariant characteristics.

[1]Shao et al. (2023) provides evidence to show that persistent absences do vary based on individual characteristics.

Even with the benefit of survey data, the analysis is still challenging due to the potential of unobserved confounding factors. Pupils who miss school may also have low motivation, a health condition, challenging home environments with poor relationships with parents or, lower quality school with less experienced teachers, which could create a spurious correlation between absences and educational outcomes.

To control for the endogeneity of absence, we exploit the panel structure of the data. Taking our baseline model from Cattan et al. (2023), we exploit within-student, between-school-year variation in absence and academic performance at three time points (year 6, year 9 and year 11).[2] We build on the baseline model by controlling for potentially time varying characteristics which we can measure using the survey data, for example parental income. We present a series of robustness tests including distinguishing between type and timing of absence, which suggest that the existence of an omitted variable bias, reverse causality or measurement error are unlikely, which in turn supports a causal interpretation.

Employing a multi-dimensional fixed effect model, we find three main results. Firstly, being absent from school has a negative and significant impact on end of year educational attainment in Maths/English. Being absent for one week leads to 3% of a standard deviation reduction. This impact is similar to studies in America and Sweden, see Aucejo & Romano (2016) and Cattan et al. (2023), whose estimates range from 3.35% to 2.75% of a standard deviation. Secondly, our findings suggest that the marginal impact of absences on both Maths and English is approximately linear. This finding is in line with Aucejo & Romano (2016), Liu et al. (2021), and Cattan et al. (2023) who also do not find evidence of non-linear impacts. Finally, we find heterogeneous impacts across pupil ability, indicating that reducing absences among lower-performing students could help close the existing performance gap.

This paper relates to a broad literature on instructional time. Finding a source of exogenous variation in instructional time is difficult, with most studies turning to variation in the length

[2] Age 11,14 and 16.

of the school year (Leuven et al. 2010, Carlsson et al. 2015, Fischer et al. 2020).

Absences are another key factor influencing the total time spent in school. This is where we focus our research. Fewer studies have used this type of variation. Four recent studies that examine the impact of pupil absences on their academic achievement are Goodman (2014), Aucejo & Romano (2016), Liu et al. (2021), and Cattan et al. (2023). Goodman (2014), Aucejo & Romano (2016) and Liu et al. (2021) all use US data. Cattan et al. (2023), who use data from Sweden, is the first study to attempt to identify the causal impact of absences using data outside the US. By applying different combinations of pupil, school, teacher, grade, and classroom fixed effects (FE) to account for the endogeneity of absences, they all provide similar estimates of negative and significant impacts of absences on achievement.

Against this backdrop, we make three contributions to the literature. Firstly, whilst we can infer answers from the few causal studies produced in the US and Sweden, the question remains as to whether these findings can be generalised to other contexts. We use England as our case study to add to the literature on the impact of absences outside the US. England provides a useful case study as comparisons can be drawn with the US and Sweden allowing for the generalisability of findings to be assessed. Secondly, we focus on absences which took place between 2006-2013, a more current sample than Cattan et al. (2023), who focuses on pupils born between 1930- 1935. Whilst this sample provides them with the benefits of examining the long run impact, it makes their short run findings out of date and makes generalisation difficult. Finally, we are the first to link administrative data on absences to survey responses. The addition of survey responses allows for richer information on individual and family characteristics. Fixed effect estimation, adopted by the majority of the literature, relies on the assumption of time invariant unobservables. The survey data allows us to assess the variability of some of variables, such as household income, which are assumed to be time invariant by the current literature.

We show that every day in school matters. This research provides evidence in support of a causal relationship between absences and achievement in England. We add to the small number of studies that examine the relationship outside of the US context, producing similar estimates, which help to determine whether findings can be generalised.

This research is organised as follows. Section 3.2 reviews the existing literature. Sections 3.3 and 3.4 outline the data and methodology respectively. Section 3.5 presents the estimated short run impact and tests the robustness, before concluding in Section 3.6.

3.2 Literature Review

3.2.1 Theory

Economists consider Becker (1962) Human Capital Theory and Spence (1973) Signalling theory to be the dominant explanations for the returns to education. The debate on which explanation should be given the most weight remains.

Human Capital Theory argues that education directly increases productivity. Signalling theory, by contrast, argues that productivity is innate, and education is just used to signal this rather than enhance it.

For our short run analysis, Becker (1962) Human Capital Theory predicts that education is positively correlated with productivity. The more time you spend in education, the higher the returns. This leads to a view that instructional time is important for pupil achievement.

3.2.2 Overview

There is a broad literature that examines the relationship between instructional time and pupil outcomes (Cooper et al. 1996, Kuhfeld et al. 2020). Whilst the studies covered in these reviews provide strong evidence of a negative correlation between absences and achievement, many of their methodologies do not permit them to estimate causal effects. This is mainly due to individuals choosing whether to attend school, but also higher quality schools offer more in person time than lower quality schools. Identifying a source of exogenous variation in instructional time is challenging, leading most studies to rely on differences in the length of the school year.

These studies leverage laws or policy changes that create exogenous variation in the length of the school year (Leuven et al. 2010, Pischke 2007, Sims 2008, Agüero & Beleche 2013, Fischer et al. 2020); differences in test dates, where the overall amount of school time remains the same but some students are tested earlier than others (Carlsson et al. 2015, Fitzpatrick et al. 2011); and unexpected school closures caused by extreme weather events (Marcotte 2007, Marcotte & Hemelt 2008, Marcotte & Hansen 2010, Hansen 2011).

Absences are another key factor influencing the total amount of time spent in school. Whilst fewer studies have used this type of variation, three recent studies that have successfully analysed the impact of absences are Goodman (2014), Aucejo & Romano (2016), and Liu et al. (2021). Although these papers are significant in advancing beyond the previous correlational evidence, they all focus on the US. Cattan et al. (2023), who use data from Sweden, was recently the first to analyse the impact of school absences on outcomes in a non-US context. Note however, that their sample covers pupils born between 1930- 1935, making their findings potentially out of date.

3.2.3 Causality problems

Pupil absences are not exogenous as individuals choose whether to attend school meaning identifying a causal relationship is a challenge.

Characteristics of pupils with high absence rates differ from those with low absence rates. Aucejo & Romano (2016) show that pupils of higher ability have lower absence rates and higher achievement compared to those of lower ability. Other factors, such as poor health, disengaged parents, and strained relationships with teachers, could create misleading correlations between absences and achievement. Not controlling for these omitted variables could lead to a downward bias on the absence coefficient (amplifying any negative causal effect) due to endogenous selection.

The literature has tried to deal with the problems of endogeneity in two ways, employing instrumental variables and panel data estimation techniques, most convincingly by Goodman (2014), Aucejo & Romano (2016), Liu et al. (2021), and Cattan et al. (2023).

3.2.4 Studies using Panel data methods

The most common methodology in the literature is fixed effects estimation.

Cattan et al. (2023) argue that the endogeneity of absences is largely driven by time-invariant factors such as ability, motivation, parental support, or school quality. Fixed effect analyses control for time-invariant characteristics and therefore, arguably correct for a lot of the endogeneity.

Studies such as Aucejo & Romano (2016), Liu et al. (2021), and Cattan et al. (2023) employ different combinations of fixed effects for pupils, schools, teachers, grades, and classrooms to address the endogeneity of absences.

Aucejo & Romano (2016) and Cattan et al. (2023) both control for pupil, teacher, and school-specific time-invariant unobservable characteristics by employing a three-way high-dimensional fixed effects model. Liu et al. (2021) control for individual and classroom fixed effects whilst additionally controlling for time-varying individual-level unobserved shocks by exploiting within-grade, between-subject variation in absences.

Aucejo & Romano (2016) and Liu et al. (2021) both use US data from North Carolina and California respectively. Aucejo & Romano (2016) analyse the impact of absences on elementary age pupils whilst Liu et al. (2021) sample covers secondary school-aged pupils. They cover a similar period, 2006-2010 and 2002-2013 respectively. Cattan et al. (2023), use data from Sweden and analyse the impact of absences on pupils who were born between 1930 and 1935.

Aucejo & Romano (2016) first estimate an OLS specification and find that the coefficients on absences for both Maths and Reading are large, negative, and significant at 0.0198 and 0.0113 of a standard deviation respectively. However, as discussed above, the value of the coefficient is likely to be biased downwards, producing more negative estimates than expected. Including individual, school, and teacher fixed effects still produces negative and significant estimates but reduces the impact on absences to 0.0067 of a standard deviation for Maths and 0.0036 for Reading. In addition, their preferred specification includes controls for peer quality and produces negative and significant coefficients of 0.0055 of a standard deviation in Maths and 0.0029 in Reading. Despite analysing absence in different countries and very different time periods, these results are similar to Cattan et al. (2023), whose OLS estimate indicates that one additional day absent leads to a 0.0055 of a standard deviation reduction in pupil performance which is measured as the average grade points across Maths, Writing, Reading and Speaking. After including fixed effects, the impact of a day absent reduces to -0.0045 of a standard deviation.

Liu et al. (2021) go further than previous studies by arguing that they have not fully addressed the issue of time-varying, pupil-specific shocks. For example, illness or a family emergency meaning previous estimates are not causal.

Liu et al. (2021) have access to subject-specific absence data, allowing them to utilise variations in absences within grades but between subjects to account for time-varying individual-level unobserved shocks. They achieve this by using the total annual absences in Maths and English Language as a proxy for year-specific unobserved shocks. They first estimate a value-added model controlling for observed pupil characteristics alongside classroom and neighbourhood-by-year FE. They estimate that one Maths absence reduces achievement by 0.00082 of a test score standard deviation. For English Language, the estimate was smaller at 0.00064 of a test score standard deviation. This finding is likely biased upwards in absolute terms due to the influence of unobserved pupil-year shocks that simultaneously affect both absences and achievement.

Their preferred model, which accounts for unobserved pupil-year shocks, reduces the point estimate for Maths, indicating that the effect of a single Maths absence is approximately -0.00042 of a test-score standard deviation. They suggest that this reduction indicates the estimate from the initial model, along with many existing estimates of the impact of pupil absences that use lagged test scores or pupil fixed effects approaches, are biased because they do not account for unobserved idiosyncratic shocks. The estimates for English Language are consistent with those for Maths.

Liu et al. (2021) findings are significantly smaller than those of Cattan et al. (2023) and Aucejo & Romano (2016). This could be expected as Liu et al. (2021) sample is of secondary-age pupils who may be more able to catch up on missed learning themselves, whilst Cattan et al. (2023) and Aucejo & Romano (2016) both analyse the impact on younger-aged children who are more dependent on others for their learning. Liu et al. (2021) also add additional controls for idiosyncratic shocks which could explain some of the difference in findings.

Cattan et al. (2023), Aucejo & Romano (2016), and Liu et al. (2021) all assume linearity. When testing this assumption, they find no evidence of non-linearities, indicating that the marginal effect of absences is roughly linear—each additional absence leads to a consistent amount of learning loss, regardless of the total number of absences a pupil has already accumulated. However, Liu et al. (2021) do state that the non-linear estimates are noisy for higher absences due to the small sample size.

3.2.5 Studies using instrumental variables

Instrumental variables correct for endogeneity if the instrument predicts changes in absences but is unrelated to changes in educational outcomes. Instruments used in the literature include pupils' distance from school (Gottfried 2010), weather-related shocks such as variation in snowfall (Goodman 2014), and infectious diseases such as flu cases (Aucejo & Romano 2016).

Gottfried (2010) analyses the impact of days present at school on test performance in urban schools. He proposes the use of distance from school, measured in mileage, as an instrument for attendance. He argues that the further away from school a pupil lives, the more absences the pupil has as commuting faces the pupils with impediments such as crowded vehicles, traffic congestion, and failure of vehicles to show up on time which can all frustrate the pupil's ability to reach school. He provides evidence to show that the correlation between distance and attendance is –0.59 for elementary and middle school pupils.

Goodman (2014) proposes the use of local variation in snowfall, defined as school days with four or more inches of snow, as an instrument for absences. The justification for using snowfall as an instrument is that pupils are more likely to be absent due to weather such as snow as it becomes dangerous to travel. Goodman (2014) provides results from a fixed-effects first stage regression to show that each excess snowy day induces a highly statistically significant 0.08 additional absences, satisfying the validity criteria.

Aucejo & Romano (2016) propose the use of a different instrumental variable, the number of flu cases. They argue that flu cases are a valid instrument because the flu is a contagious respiratory illness that impacts all age groups, with school-aged children experiencing the highest rates of infection. Their first-stage results indicate a positive and significant relationship between flu cases per 10,000 school-aged children and absences, meeting the validity criteria.

Gottfried (2010), Goodman (2014), and Aucejo & Romano (2016) all use US data covering three areas, Philadelphia, Massachusetts, and North Carolina respectively. Gottfried (2010) uses data from 1994-2001, Goodman (2014) from 2003-2010, and Aucejo & Romano (2016) use data covering the period 2009 to 2010.

Goodman (2014) results suggest that each day missed due to bad weather reduces Maths scores by a marginally significant 0.023 of a test score standard deviation. The impact on English Language is small (-0.004) and not statistically significant. Aucejo & Romano (2016) find slightly different results to Goodman (2014) by showing that being absent has a negative and statistically significant impact on Reading test scores of 0.018 of a standard deviation whilst for Maths, they find a statically insignificant effect. Gottfried (2010) estimates that a one standard deviation increase in the days a pupil is present in school is associated with a statistically significant 0.0045 for elementary and 0.0039 for middle school pupils, standard deviation change in the grade point average (GPA).

Aucejo & Romano (2016) IV estimates are almost five times larger than their OLS estimates, (0.0182 compared to 0.0043). Even though Gottfried (2010) IV estimates are the smallest in the literature, he still finds results from instrumental variable estimation that are larger than OLS (0.0045 compared to 0.0024). Further discussion of this is provided below.

Goodman (2014) expands his analysis by demonstrating that absences are not the only way weather impacts instructional time. He finds that each additional day with four or more inches of snow leads to a highly statistically significant 0.09 increase in school closures. Consequently, snowy days affect instructional time through two mechanisms—absences and school closures—thereby violating the exclusion restriction. Therefore, using snowfall alone cannot determine whether the negative effects on Maths achievement are due to school closures, pupil absences, or a combination of both types of lost instructional time.

Goodman (2014) introduces another instrument—the number of days with 10 or more inches of snow—which changes the instrumental variable estimates of the impact of lost instructional time. Each additional extremely snowy day leads to a substantial increase in school closures but does not lead to more absences. As a result, Goodman (2014) uses moderate snow, defined as days with 4-10 inches of snowfall, to instrument for absences, and extreme snow, defined as days with 10 or more inches of snowfall, to instrument for closures.

The findings now show that each additional absence decreases Maths achievement by a large and statistically significant 0.05 standard deviations, whilst the point estimate on the effect of school closures is almost identical to zero. The findings for the impact on English Language are negative but statistically insignificant. These findings are considerably larger than his findings when using just one instrument.

Goodman (2014) suggests that the difference between the effects of absences and closures could be due coordination of pupils set out in the model of instruction. When a pupil returns to school following a period of absence, they require differential support in order to access work they have missed. However, when the whole school is closed, it is easier for the teacher to compensate for the missed learning at no detriment to the pupils.

3.2.6 Heterogeneity

The impact of absences may vary depending on pupils' characteristics, with catching up after an absence likely being more challenging for low-performing students. While much of the literature suggests that absences are equally detrimental across different demographic groups, school types, and grade levels (Liu et al. 2021, Cattan et al. 2023), some studies have shown that low-income pupils are more negatively affected by school absences than their wealthier counterparts (Aucejo & Romano 2016, Gershenson et al. 2017). Additionally, Aucejo & Romano (2016) find that students in the lowest tercile of prior academic performance are the most adversely impacted by an additional absence.

3.2.7 Critical analysis

Despite methodological improvements such as controlling for time-invariant observable and unobservable confounders or identifying exogenous variation in absences due to random events, weaknesses still exist. The requirements for a valid instrument are difficult to meet and not all confounders are time-invariant.

Validity of instruments

The robustness of observational studies such as Gottfried (2010), Goodman (2014), and Aucejo & Romano (2016) hinges on the strength of their instrumental variables. However, this is not straightforward, raising concerns that endogeneity issues may not have been fully addressed. Although these studies consistently find significant negative effects of absences on achievement, ranging from 0.45% to 5% of a standard deviation, the validity of the relevant exclusion restrictions remains open to debate.

IV strategies that rely on weather or flu cases might also influence other pupils and teachers in the class, meaning that estimates derived from these instruments will not accurately measure the impact of individual school absences while holding all other factors constant. Whilst Aucejo & Romano (2016) recognise the limitations of their instrument and control for peer absences, they do not have data on teacher absences for the whole of their research time period. However, using 2008 teacher absences due to sickness as a proxy, they find that adding this control to the model does not affect the coefficient on days absent and does not substantially affect pupils' performance. Goodman (2014) also lacks data on teacher absences but argues that this is not a significant issue. Research indicates that weather generally does not have a strong impact on teacher absences (Herrmann & Rockoff 2012), and even if it did, the effect on students' achievement would be too minor to account for the estimated impact of absences on students' achievement (Fryer 2013).

Another reason flu cases may not be exogenous is that they may directly impact test scores if a pupil is ill on the day of the test. However, Aucejo & Romano (2016) present data showing that the flu season peaks in January, February, and March, which is well before the tests are administered in the summer months, with flu cases during the exam period being very low.

Despite Goodman (2014) acknowledging the non-linearity of the instrument[3], it could still not be exogenous. In Massachusetts, where snow is a frequent winter occurrence, schools might already account for weather-related absences when planning their annual school calendar.

Also, instruments such as snow and distance from school could be correlated with tardiness[4] which is not included in the Goodman (2014) or Gottfried (2010) models. Whitney & Liu (2017) look at the impact of half-day absences on pupil achievement and found that not including half-day absences or tardiness in the model could lead to understating the true impact of absences.

It is also possible that distance from school could be associated with family characteristics, as parents often choose neighborhoods based on their preferences. However, Gottfried (2010) counters this by presenting correlation coefficients and their significance levels between distance and variables such as GPA, lagged GPA, and other pupil and neighborhood

[3] As snow fall increases, this leads to school closures rather than absences meaning more pupils are absent from school.

[4] Living far away from school and snow fall could both cause problems when travelling to school.

characteristics. The results show that, for all variables, the correlations are extremely low and statistically insignificant.

Gottfried (2010) also performs a robustness check by reducing the sample size to pupils who remained living at the same location but moved from middle school to high school. While families may choose where to live based on middle school locations, if they did not move when the pupil moved to high school then we can assume that they did not choose where to live based on high school location. He finds that a one standard deviation increase in the days a pupil is present in school is associated with a statistically significant 0.0053 standard deviation change in GPA. This is slightly higher than the previous results, though, the result is still positive and statistically significant, supporting the hypothesis that there is a positive payoff from attending school.

Omitted variables

The main challenge to identification in panel data estimation is the possibility that unobserved, time-varying shocks, such as illness or family emergencies, could confound the estimates by affecting both pupils' attendance and their academic performance.

Liu et al. (2021) specifically control for family shocks, finding estimates smaller than studies such as Aucejo & Romano (2016) who did not. Aucejo & Romano (2016) estimate a family-year fixed effects specification by comparing siblings and find estimates that are similar to their original estimates, which suggests that family-specific shocks are not driving their results.

Additionally, absences are primarily caused by illness[5]. Consequently, a concern is that an idiosyncratic health shock could confound the impact of lost instructional time. Aucejo & Romano (2016) run a sensitivity test by splitting absences into excused, which are more likely

[5] Illness accounted for 52.6% of all absences in 2018/19 (Department for Education 2019).

to be illness-related, and unexcused. If health shocks were influencing the results, we would expect excused absences to have a more detrimental effect compared to unexcused absences after disaggregating the types of absences. However, both specifications show similar results. An additional excused absence decreases Maths (reading) scores by 0.41% (0.18%) of a standard deviation, while unexcused absences reduce scores by 0.7% (0.4%) of a standard deviation. These findings suggest that health issues do not appear to bias the results. This conclusion is supported by Cattan et al. (2023), who also find no significant difference in the marginal effects of absences due to sickness compared to those due to other reasons.

Pupil-fixed effects control for time-invariant pupil disengagement. However, since pupils may lose interest in classroom activities over time, the dynamic aspect of disengagement cannot be captured by just including pupil-fixed effects. This raises the concern that the observed effects of absences might be influenced by a correlation between "disengagement from school" and the decision to miss school. To address this, Aucejo & Romano (2016) re-run their analysis including total days suspended as a proxy for pupil disengagement. If disengagement were impacting the results, we would expect the effect of absences to decrease once suspensions are accounted for. Their results show that the coefficients on absences remain fairly stable, lending credibility to their findings.

Similarly, Liu et al. (2021) uses a proxy identification strategy that assumes pupils' preferences for different subjects do not change endogenously over time. This assumption would not hold true if, for instance, a pupil's improved performance and reduced absences in a class were due to having friends in that class or a strong connection with the teacher. To validate this assumption, Liu et al. (2021) conduct an auxiliary analysis showing that while Maths and English Language absences in a given year are highly correlated (0.85), two-thirds of pupil-year observations have different numbers of absences in each subject. They further demonstrate that absences in both subjects respond similarly to lagged test scores and prior absences in Maths. If subject-specific preferences or abilities were driving the differences in absences, we would expect English Language absences to be less affected by Maths performance.

Generalisability of findings

The literature is consistent in finding negative impacts of absences on achievement. However, depending on the estimation strategy used, the magnitude can vary considerably. Aucejo & Romano (2016), Liu et al. (2021), and Cattan et al. (2023) using fixed-effect estimation find estimates smaller than OLS. Gottfried (2010), Goodman (2014), and Aucejo & Romano (2016) all using instruments, find estimates significantly larger than their OLS estimates. The instrumental variables estimates of the impact of absences are approximately two and a half times larger than the fixed effects estimates. Given the numerous omitted variables in the OLS model, we would anticipate that OLS estimates are biased downwards, resulting in more negative estimates than would be expected. Fixed effect estimation controls for some of these omitted variables, producing less negative estimates. Instrumental variable methods produce LATE estimates, which could explain the larger effects found.

Goodman (2014) provides the largest estimates in the literature, finding that ten days of absences due to bad weather reduces Maths achievement by 50% of a standard deviation. In contrast, Aucejo & Romano (2016) report much smaller estimates, ranging from 5% to 10% of a standard deviation, depending on the specification. The discrepancies between these results may stem from several factors. Firstly, the number of fixed effects included in the models may differ. Secondly, the use of different instruments could lead to variations in local average treatment effects. For instance, if snowfall increases absence rates more for disadvantaged students, the marginal pupils identified by the instrumental variables may be disproportionately disadvantaged. Since disadvantaged students may benefit more from school attendance, the IV estimates could be larger. Finally, fixed effects estimates might be slightly biased downward because the absence measure includes the entire year, including post-exam absences. Goodman (2014) suggests that if absences are assumed to be evenly distributed throughout the year, full-year absences might overstate annual fluctuations in pre-exam absences by about 10% in Maths and 30% in English Language. In contrast, Goodman (2014)'s instrumental variables estimates do not suffer from this issue, as all snowfall—and thus all snowfall-induced absences contributing to identification—occurs before exams.

3.2.8 This research's contribution

Analysing the relationship between pupil absences and outcomes is impacted by methodological challenges. Research has varied considerably in terms of methodologies and time spans, generating similar findings. In 16 studies over 14 years, research has consistently found that absences have negative effects on test scores. However, there have been large inconsistencies in the magnitude, results have ranged from finding no statistically significant relationship for certain subjects, to large negative relationships ranging from -0.45% of a standard deviation to -5%. Little is currently known about the underlying sources of this variability as the majority of the literature suggests that absences are similarly harmful to all pupils. Therefore, it is important to carry on updating the analysis using a range of countries.

This research contributes to the literature by analysing the impact of school absences on pupil outcomes in England. To the best of my knowledge, there are only two other studies, Department for Education (2016) and Arulampalam et al. (2012), that look at the impact of instructional time on achievement in England. Department for Education (2016), using data from the National Pupil Database, looks at the correlation between pupil-level absences and achievement at the end of KS2 and KS4. Whilst they have the benefit of a large sample size, they are unable to reach causality due to only having data on a small number of covariates. Arulampalam et al. (2012), using the random assignment of students to classes on different days and times of the week as an instrument for absences, focus on students from a single English university. As discussed, the use of instrumental variables generates local average treatment effects which can mean that the estimates found are sometimes difficult to generalise.

As discussed in Section 3.1, we make three contributions to the literature. Firstly, we use England as our case study to add to the literature on the impact of absences outside the US. Secondly, we focus on absences which took place between 2006-2013, a more current sample than Cattan et al. (2023), who focus on pupils born between 1930- 1935. Finally, we are the first to link administrative data on absences to survey responses. The addition of survey responses allows for richer information on individual and family characteristics.

3.3 Data

3.3.1 Data Source

The data used in this research combines the National Pupil Database (NPD) with Understanding Society (UKHLS)- Harmonised British Household Panel data (BHPS).[6]

The NPD is an individual-level administrative database held by the Department for Education in England.[7] The database combines information held by schools, exam awarding bodies, and local authorities on all pupils, aged 2-18, in English state schools. In 2019/20 8.3 million pupils were recorded in the NPD (Department for Education 2022B). The NPD is a

[6]Data Citation: University of Essex, Institute for Social and Economic Research. 2022. Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009: Secure Access, [data collection]. 13th Edition. UK Data Service. SN: 6676, DOI: 10.5255/UKDA-SN-6676-13 Department for Education, University of Essex, Institute for Social and Economic Research. 2022. Understanding Society: Linked Education Administrative Datasets (National Pupil Database), England, 1995-2018: Secure Access. [data collection]. 3rd Edition. UK Data Service. SN: 7642, DOI: 10.5255/UKDA-SN-7642-3

[7]The NPD was first produced in 2002 and is mainly used for funding purposes, school performance tables, policy making, and research.

valuable data resource, offering an almost complete view of school outcomes for the majority of children in England. However, it does not include children who attend private schools or those who are home schooled.

The NPD is made up of multiple data sets which contain a range information on pupils' education attainment at different Key Stages. The School Census, conducted three times a year (in January, May, and October), collects data on gender, ethnicity, first language, eligibility for free school meals, special educational needs, and records of absences and exclusions. The NPD also contains information on pupils' test results across all Key Stages. In Key Stage 1, tests are teacher-assessed for English, Mathematics, and Science. In Key Stage 2, pupils take tests in the same subjects as in Key Stage 1, with both the level achieved and the point score being recorded. For Key Stage 3, the level achieved in English, Mathematics, and Science tests is available. Achievement data in Key Stage 4 includes GCSEs and equivalent qualifications.[8] The grade achieved in A levels and equivalent qualifications is reported for those individuals who complete post-compulsory education.

The second dataset used in this research is Understanding Society- Harmonised British Household Panel data (UKHLS). The harmonised data set is comprised of the BHPS which ran from 1991 to 2009 and Understanding Society which took over in 2009, growing the sample to cover over 40,000 households[9].

UKHLS surveys a sample of individuals representing the UK population[10]. It is a longitudinal survey covering a multitude of topics. One household member responds to the household

[8]Equivalent qualifications include vocational qualifications such as BTECs.

[9]The BHPS was a longitudinal social survey of households and individuals in the UK. The survey covered 10,000 households and ran from 1991 to 2009. Data collection under the umbrella of the BHPS study title stopped in the year Understanding Society started (2009). Understanding Society was built on the success of BHPS, covering the same core questions whilst building on the BHPS original sample. This offers an opportunity to exploit data from the two studies jointly to create a long panel of data. The 'Understanding Society harmonised BHPS' project started in 2016.

[10]The sample is large enough to have around 10,000 people for each birth cohort per decade from the 1940s onwards. They also have approximately 17,000 children who have been born into the Study since the year 2000, over 18,500 two generation and 2,700 three generation families.

questionnaire which covers household amenities, expenses, rent/mortgages, deprivation, etc. There are also individual questionnaires, children aged 10-15 complete the youth survey whilst everyone 16+ responds to the adult survey. Some information about 0–9 year-olds is collected from their parents/guardians. Basic demographic information is collected alongside key topics ranging from education and employment to political attitudes and health behaviours. Topics covered in the Youth survey include wellbeing, relationships with family and friends, views on school and further aspirations, etc.

The NPD can be linked to UKHLS using pupil-level identifiers. Consent for data linkage was obtained for both adults young enough to have educational records in the NPD and for children aged 4-15.[11] Individuals aged 16 and older could provide consent on their own, while responsible adults gave consent on behalf of children under 16. In Understanding Society, consent for education data linkage is gathered every three years, with the first round collected in Wave 1 (2009) and the second in Wave 4 (2012). In both waves, the consent process involved distributing an information leaflet to participants, asking a consent question in the main survey, and obtaining a signed consent form. Potential sample selection bias is discussed in Section 3.5.4.

The main benefit of this data linkage is it allows the individuals to be matched to their survey responses and their parents' survey responses. This allows individual and family characteristics to be controlled. In the literature to date, family characteristics are assumed to be constant over time and therefore controlled via individual fixed effects. However, family income, for example, is likely to vary over time. If family income impacts both pupil absence and achievement, then the estimation will be biased. This research is able to control for this confounder outside of the fixed effect.

[11]NPD data is available for individuals born after 1985.

3.3.2 Sample

The sample is made up of 4,088 individuals across a minimum of two key stages, creating a total sample of 9,279 observations.[12] This analysis is restricted to the years 2006-2013 and Key Stage 2-4.[13]

Whilst the NPD has close to population coverage, the sample used in this research is constrained to the number of individuals who are in Understanding Society and are willing and able be matched.[14] Matching is currently complete for Wave 1 and Wave 4 of Understanding Society.

Absences were first recorded at the pupil level in the 2005/6 academic year.[15] From then on, absences are recorded in every year from Reception to Year 11. Post-KS4 absences are not recorded as education was not compulsory post-16 until 2013.[16] Even when the education participation age was changed, KS5 attendance data is not recorded in NPD as attendance requirements differ by what pupils do at age 16-18. Some pupils will go on to sixth form whilst others will undertake apprenticeships. Absences are recorded in KS1, though as no one in this sample has absence data recorded in KS1, this key stage is not included in this research sample.[17]

[12]This is an unbalanced panel. There are two reasons why some individuals do not have data in all three key stages. Firstly, some are missing KS2 data because they completed KS2 prior to the 2006/7 academic year and therefore NPD does not record individual level absences. Secondly, they are missing KS4 data because they are yet to reach this level or the data is yet to be matched.

[13]The analysis is restricted to between 2006-2013 as only Waves 1-4 of Understanding Society have been linked to NPD at the time of the analysis.

[14]Figure B.1 in the appendix shows the proportion of individuals eligible, willing, and able to be matched.

[15]In 2005/6 absences were recorded for primary school pupils only. In the 2006/7 academic year, absences for both primary and secondary pupils were recorded.

[16]In September 2013 the education leaving age was raised to 17, and from September 2015 it was raised to 18.

[17]Absence data started being recorded at an individual level is 2005/6. Since the sample period covers individuals in KS2 from 2006-2013, no one in the sample has KS1 absence data.

3.3.3 Outcome variable

This research measures educational achievement by Maths and English test scores which are recorded in the NPD by Key Stage. It is compulsory to take Maths and English exams in the last year of every Key Stage. Science is not included in this analysis as compulsory science exams were discontinued for KS2 and KS3 in 2010.

In May of year 6 (age 11) KS2 Standard Assessment Tests (SATs) are taken. These are formal tests in English (grammar, punctuation, spelling, and reading) and Maths. All pupils, across England, take the same exam on the same day. SATs are both set and marked externally.

Prior to 2008, pupils in state-funded schools also took KS3 SATs in year 9 (age 14). However, KS3 SATs were abolished in 2008. Pupils are now assessed by their teachers. Pupils' achievement is assessed relative to the national curriculum programmes of study. National curriculum levels range from 1-8 with the expected level achieved at the end of KS3 a 5/6. National curriculum levels achieved for Maths and English assessments are recorded in the NPD. [18]

At the end of KS4 (year 11), pupils aged 16 take their General Certificate of Secondary Education (GCSE) exams. The total point score for all GCSE and equivalent qualifications taken by the pupil is recorded in the NPD. The breakdown by subject is given by grade achieved. All pupils must take Maths and English, though they do not all take the same exam as there are a range of exam boards overseen by Ofqual. At the time of the data, GCSE exams are graded between A* (the highest) and G (the lowest). For the main analyses, the grade achieved in Maths and English is used to construct the dependent variable .[19]

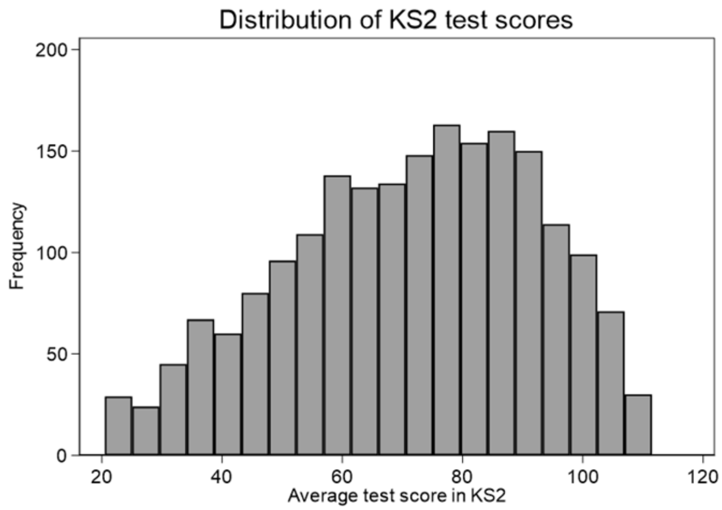
Figures 3.1, 3.2, and 3.3 illustrate the distribution of the average point score for Maths

[18] All KS3 test scores in this research are teacher assessed. In Section 3.5.4, the sample is restricted to those pupils with externally marked KS3 exam results prior to 2008 as a robustness check.

[19] In Section 3.5.4, the point score across all GCSE and equivalent qualification is used as an alternative dependent variable.

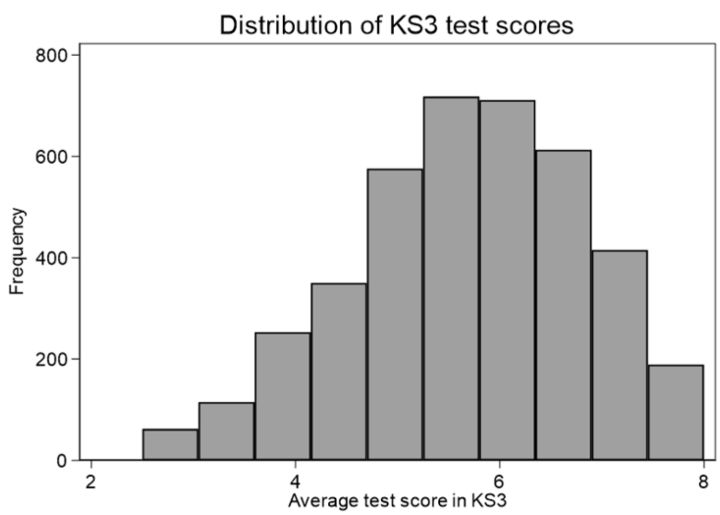
and English in each Key Stage. All test score distributions are normal with a few pupils receiving very low or very high grades. Appendix Figures B.8- B.13 show the distributions of grade points by subject and Key Stage.

Figure 3.1: Distribution of KS2 test scores in Maths and English



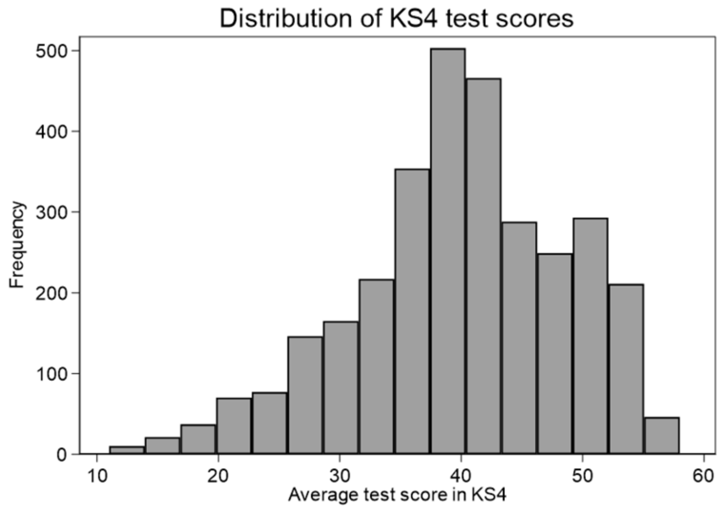
Note: The data covers the whole sample period from 2006-2013. The point score is taken from the average achievement in Maths and English.

Figure 3.2: Distribution of KS3 grades in Maths and English



Note: The data covers the sample period from 2008-2013. The grade is taken from the average achievement in Maths and English. KS3 achievement is measured in National Curriculum levels.

Figure 3.3: Distribution of KS4 test scores in Maths and English



Note: The data covers the whole sample period from 2006-2013. The test score is taken from the average achievement in Maths and English. KS4 achievement in Maths and English is measured by grades. Each grade is allocated an amount of points. 40 is a grade C which is defined as a pass.

To enable comparisons and control for the difference in assessments taken, test scores are standardised within Key Stages to have a mean of 0 and a standard deviation of 1.

3.3.4 Absences

The main explanatory variable in this research is pupil absences.

All public schools must offer two sessions per day, one in the morning and one in the afternoon. The duration of each session, as well as breaks and the overall school day, is set by the school's governing body. However, schools are required to provide a total of 380 sessions or 190 days within each academic year.

Schools must record attendance twice daily: once at the start of the morning session and once during the afternoon session. Consequently, absence data is logged by session. Schools are required to note whether a pupil is present, participating in an approved educational activity, absent, or unable to attend due to exceptional circumstances. When a pupil is absent, schools also must determine the reason and indicate in their register whether the absence is authorised or unauthorised. Authorised absences are where the school has received an acceptable explanation for the absences such as illness, appointment or religious observation. An unauthorised absence is where no such explanation has been given or the school has denied permission for the absence such as family holidays. Beyond the classification of absences as authorised or unauthorised, there is limited information available about the specific reasons for the absence. [20]

Pupil level absence data is aggregated over a term. The NPD does not collect dates of absence and therefore, this research is unable to identify the length of individual absence spells.

3.3.5 Descriptive statistics

Table 3.1 provides descriptive statistics on the sample of pupils analysed in this research. Pupils are absent on average for 18.6 sessions per year. 50.9% of the sample are female. 62.9% are white, 18.5% are asian and 9.6% are black. 9.9% have a special educational need and 11.6% are eligible for free school meals. This is representative of the English population.

[20] 4.6% of the sample report the reason for absence.

Table 3.1: Summary statistics

Variable	Mean
Absence	18.57
Gender(%)	
Female	50.95
Male	49.05
Ethnicity (%)	
White	62.88
Black	9.55
Asian	18.46
Mixed	7.48
Other	1.63
Other characteristics (%)	
Special Educational Need	9.95
Free school meal eligibility	11.59
Highest qualification of parents is a degree or equivalent	19.59
Highest qualification of parents is other higher or equivalent	15.45
Highest qualification of parents is A-levels or equivalent	16.97
Highest qualification of parents is GCSEs or equivalent	27.93
Highest qualification of parents is other qualification	9.80
Parents have no formal qualifications	10.26
Monthly household income	3,109

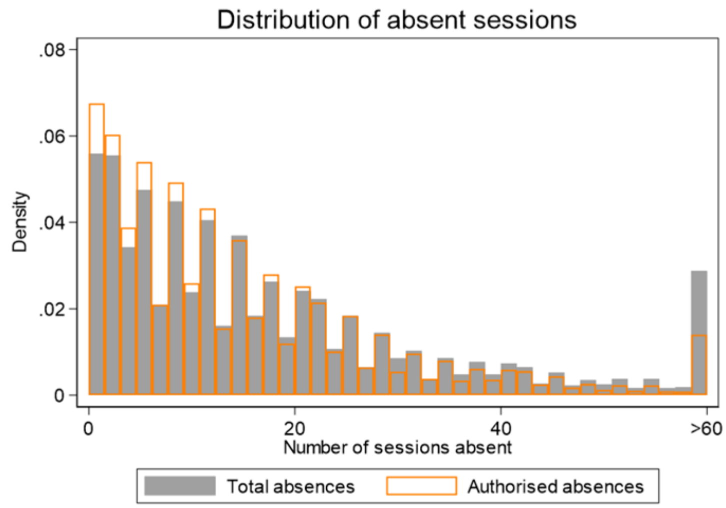
Figure 3.4 presents the distribution of total (authorised) absent sessions across all years. While the median number of sessions absent is 13 (6.5 days), a significant number of pupils are absent for extended periods. Only 6.3% of pupils have 100% attendance and 25% of pupils are absent for more than 24 sessions. Despite using data from different countries and in different time periods, the distribution of absences found here is similar to that found by Liu et al. (2021) and Aucejo & Romano (2016) in America and Cattan et al. (2023) in Sweden.[21]

Figure 3.4 illustrates that most absences are authorised. The average number of sessions missed for unauthorised reasons is 3.4 and 66.9% of pupils have never missed a session for an unauthorised reason. Appendix Figures B.2- B.4 show the distribution of total (authorised)

[21]Liu et al. (2021), Aucejo & Romano (2016) and Cattan et al. (2023) find average absence at 11.1, 6.28 and 11 days respectively.

absent sessions by Key Stage.[22]

Figure 3.4: Distribution of (authorised) absent sessions



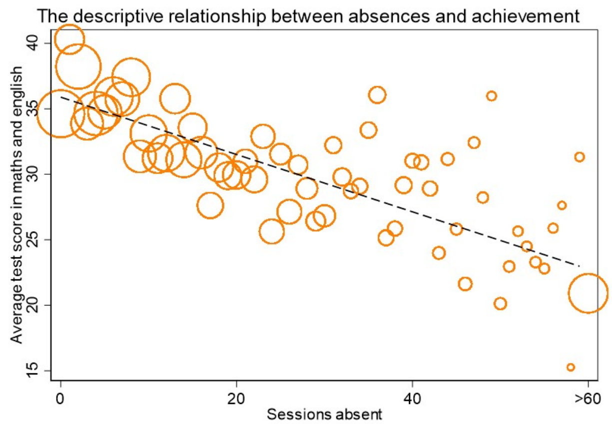
Note: Observations are counts of yearly absent sessions across year 6, year 9 and year 11.

To conclude this section, scatter plots document the association between the number of sessions absent (across all Key Stages) and the main outcomes of interest. The correlation between the number of absent sessions and average achievement in Maths and English is negative, as shown in Figure 3.5.[23] Figure 3.6 documents the correlation between authorised absent sessions and average achievement in Maths and English. Authorised absences are largely driven by illness. Assuming that illness is exogenous, the correlation between authorised absent sessions and achievement may provide more convincing evidence of a causal relationship. Appendix Figures B.5- B.7 show the correlations by Key Stage.

[22]The median number of absent sessions for KS2, KS3 and KS4 are 10,13 and 14 respectively.

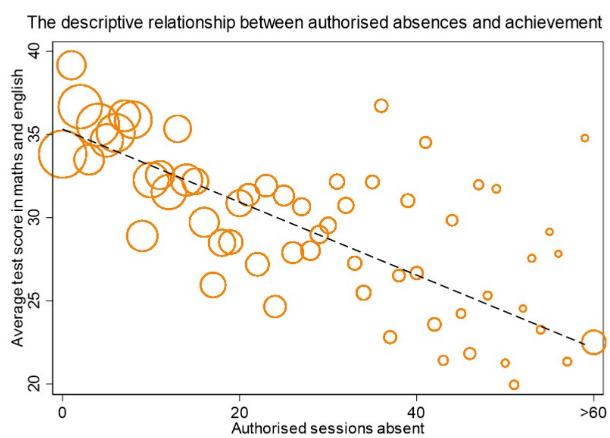
[23]For zero absences, the average test score is 35. For 10 absent sessions or a week absent, the average test score is 33 and for 20 it is 30.

Figure 3.5: Correlation between absent sessions and average achievement in Maths and English



Note: Correlation between absent sessions and average achievement in maths and English across KS2, KS3 and KS4. Test scores are collapsed on the integer of sessions absent. The size of the marker indicates the relative number of observations in the sessions-absent cell. The fitted line is taken from a simple linear regression of achievement in maths and English on total sessions absent.

Figure 3.6: Correlation between authorised absent sessions and average achievement in Maths and English



Note: Correlation between authorised absent sessions and average achievement in maths and English across KS2, KS3 and KS4. Test scores are collapsed on the integer of sessions absent. The size of the marker indicates the relative number of observations in the authorised-sessions-absent cell. The fitted line is taken from a simple linear regression of achievement in maths and English on total authorised sessions absent.

Although these correlations suggest a possible adverse impact of school absences on short-term outcomes, individuals choose whether to attend school, meaning the evidence presented above may just be correlations and not causal. Characteristics of pupils with high absence rates may differ from those with low absence rates. To start exploring the extent of such selection, Table 3.2 shows the average level of absence by individual observable characteristics. The main difference is across ability which is measured by achievement in Maths and English in the previous Key Stage. Though there is limited evidence of selection on observables, this does not rule out selection on unobservables.

Table 3.2: Average level of absence by individual characteristics

Variable	Average absence level	SD
Average absence	18.6	20.7
Male	17.8	20.5
Female	19.1	20.9
White	19.2	21.3
Asian	15.3	17.2
Black	11.9	13.4
Mixed	20.7	23.3
Other	14.5	16.9
2006	19.6	18.0
2007	16.8	18.0
2008	17.7	18.8
2009	18.2	18.8
2010	20.1	22.7
2011	19.4	21.9
2012	18.4	23.2
2013	17.9	22.7
Ability distribution		
Lowest quartile	35.0	36.0
Second quartile	19.3	19.3
Third quartile	17.7	18.2
Highest quartile	13.1	12.5

3.4 Method

The data allows this research to observe individual test scores in both Maths and English, as well as the number of sessions absent in the last year of KS2, KS3, and KS4. This analysis aims to identify the causal effect of school absences in an academic year on the achievement in Maths and English that is measured at the end of the same school year. As previously discussed in Section 3.2, establishing causality between absences and academic attainment faces methodological challenges. There are three primary threats to discerning causality: (i) omitted variable bias, (ii) reverse causality, and (iii) measurement error. This section will outline how this research will try to correct for omitted variables. The second and third threats to establishing causality will be addressed in Section 3.5.4.

This analysis starts by specifying a value-added model of the education production function where pupil absences are a contemporaneous input that relate linearly to pupil attainment.[24]

The model is specified as:

$$Y_{isy} = \beta_0 + \beta_1 A_{isy} + \beta_2 X_{isy} + \epsilon_{isy} \tag{3.1}$$

Y_{isy} is the attainment in Maths/English of pupil i in school s and school year y where attainment is standardised by subject and key stage. A_{isy} is the number of sessions of school the pupil has missed in school year y. X_{isy} is comprised of individual and household characteristics, including lagged achievement, gender, ethnicity, the month of birth, age, and two dummy variables which equal 1 if the pupil has a special educational need and if they are eligible for free school meals, respectively. Household characteristics include parents' education, which is defined as the highest education category held by either parent, and the net monthly household income.[25] ϵ_{isy} captures the unobservable determinants of achievement.

To interpret β_1 as the causal effect of a session absent on a pupil's end of year test score, we

[24]Non-linear effects are explored in Section 3.5.2.

[25]Some data on household characteristics are missing. To allow for use of all observations, this research employs the dummy variable adjustment method.

require independent variation in absences, meaning the zero conditional mean assumption must hold, $E(\epsilon_{ifys}|A_{ifys}, X_{ifys}) = 0$. Due to the endogeneity of absences, it can be argued that this assumption may not hold. Pupils and schools with high absence rates differ from those with low absence rates in both observed and unobserved aspects. As shown in Table 3.2, pupils at the higher end of the achievement distribution are more likely to attend school and score highly in their exams. Some aspects in which absence rates differ, such as lagged ability, have suitable proxies in the data. Other dimensions, such as pupil motivation, do not. Therefore, OLS estimation is likely to yield an overestimation due to the downward bias caused by endogenous selection.

To expand on the identification problem, this research will follow Cattan et al. (2023) by decomposing the error term into four sets of factors and an i.i.d. error.

$$\epsilon_{isy} = \lambda_i + \alpha_f + \rho_s + \eta_t + \mu_{isy} \tag{3.2}$$

Where λ_i is a pupil-specific unobserved factor, α_f is a family-level unobserved factor, ρ_s is a school-level unobserved factor, η_t is a teacher-level unobserved factor, and μ_{isy} is an idiosyncratic term varying across pupils and school years.

It is possible to think of unobservable factors that could simultaneously impact pupil achievement and absence. Starting at pupil-level, personality, mental health, and motivation are all unobserved within the data but likely to impact both absence rates and pupil achievement.

Family level unobservable factors could include parental interest in their children's education or parental ability to help their children, which includes educational ability and also time.

At a school level, absence policies, curriculum, hiring practices, school neighbourhood, and the quality of leadership all vary between schools. Finally, teacher quality may impact both pupil achievement and absence rates.

There may also be school year effects, for example, stress and pressure could increase in higher school years, parents could invest more time into their child at a younger age or there may be a weather shock in one year.

To try to control for these sources of endogeneity, the main estimation strategy used in this research exploits the panel properties of the data. Incorporating fixed effects removes the endogenous selection bias caused by time-invariant unobserved factors. Fixed effect estimation may also be more relevant for policy development than other quasi-experimental frameworks such as instrumental variables. This is because fixed effect models estimate the average effect on school achievement for those whose absences change over time rather than a narrowly defined local average treatment effect.

Since we track pupil achievement and absences over several school years, we can account for individual-specific, time-invariant unobservable factors by incorporating individual fixed effects. This approach helps control for constant individual and family characteristics. Additionally, we can address time-invariant unobservable factors specific to each school by including school fixed effects. Finally, we include dummy variables for both the school year and the calendar year.

Data on teachers that can be linked to the pupils they teach is currently not available in England. This research is therefore unable to control for teacher level endogeneity. Using data from America, whilst Chetty et al. (2014) finds evidence to suggest that teacher quality impacts pupil achievement, Fryer (2013) argues that teacher quality does not impact pupil absence. Gershenson et al. (2017) does argue that teachers have a causal impact on pupils' absences, though he also finds that teachers who improve test scores do not necessarily improve pupil attendance.

This identification strategy is similar to that of Cattan et al. (2023) and Aucejo & Romano (2016). This research's estimation strategy can be summarised by the following estimated equation:

$$Y_{isy} = \alpha_0 + \alpha_1 A_{isy} + \alpha_2 X_{isy} + \gamma_i + S_s + T_t + G_y + \mu_{isy} \tag{3.3}$$

Where X_{isy} controls for observable time-varying individual and household characteristics, including special educational needs, free school meal eligibility, and household income. γ_i is the individual fixed effect, S_s is the school fixed effect, and G_y and T_t are dummy variables for the school year and calendar year.

To interpret α_1 as the causal effect of a session absent on a pupil's end of year test score, we must assume that the decision to be absent is predominantly influenced by time-invariant characteristics. If there are unobserved factors which impact both absences and achievement that vary over time, the estimate will still be biased. Robustness checks are carried out in Section 3.5.4, where we disaggregate absences by type and time to check for health shocks and measurement error. We also check for reverse casualty by using the lag of absence.

3.5 Results

3.5.1 Baseline results

Table 3.3 (3.4) contains the point estimates for Maths (English) from the education production function specified in equations (3.1) and (3.3). For comparison purposes, the first three columns for Maths and English show the results when the model is estimated without fixed effects, these are referred to as pooled OLS models. Estimates are reported without ability and family controls (specification 1) and with family controls (specification 2) and then also ability controls[26] (specification 3). Specification 4 includes individual fixed effects

[26] Achievement in Maths/English in the previous Key Stage is used as a proxy for ability.

whilst specification 5 additionally controls for school fixed effects. In specification 6, we substitute individual fixed effects for lagged achievement .[27] Angrist & Pischke (2009) argue that specification (5) can be seen as a lower bound whilst specification (6) is the upper bound[28]. Achievement is standardised within-subject and key stage and all specifications assume linearity.[29]

Specification 1, estimated by pooled OLS, generates coefficients on absences that are negative, significant, and large in magnitude for both Maths and English. Since there are no controls for individual ability or family characteristics as well as unobserved factors, which are all likely to be correlated with opposite signs with absences and test scores, the OLS coefficient is likely to be biased downwards. We anticipate that, with the inclusion of appropriate controls, the magnitude of the coefficient will reduce.

Confounders such as family characteristics and ability are to some degree observable in the data. Once we control for family characteristics, including parents' education and net monthly income in specification 2, the size of the impact of absences on achievement in Maths and English reduces from 1.3% of a standard deviation for Maths and 1.1% for English to 0.97% for Maths and 0.87% for English.[30] Extending the model further to include lagged achievement in specification (3)[31], the impact of a session absent remains negative and significant though does reduce further to 0.63% for Maths and 0.59% for English.

There are still potentially unobserved time-invariant factors such as motivation, parental

[27]This research does not include pupil fixed effects and lagged achievement in the same equation due to the Nickel bias.

[28]The fixed effects model provides a lower bound because it controls for unobserved, time-invariant factors, making the estimate more conservative. The model with a lagged dependent variable provides an upper bound because it can overestimate the effect by capturing persistent unobserved factors that influence both the lagged and current dependent variable.

[29]This assumption is tested in Section 3.5.2.

[30]The literature to date has assumed that household income is time invariant and therefore have used individual fixed effects to control for this. We test this assumption and find that household income does vary over time however, this does not impact the estimated coefficient and therefore we argue that the use of individual fixed effects is sufficient to control for household characteristics.

[31]The sample size decreases as there is no measure of lagged achievement for KS2.

input, and school quality that are not controlled for in specification (3). Utilising the panel structure of the data, specification (4) includes pupil-fixed effects.[32] An additional session absent is associated with Maths and English scores declining by 0.42% of a standard deviation.

The inclusion of school fixed effects in specification (5) which controls for any observed or unobserved factors at the school level that are constant over time further reduces the estimated coefficient to 0.34% for Maths and 0.36% for English. A pupil with the median number of absent sessions, 13, would see their test scores decline by 4.4% of standard deviation in Maths and 4.7% for English, compared to a pupil with no absences.[33]

Specification (6) examines a model where we substitute pupil fixed effects for lagged achievement. Specification (5) can be seen as a lower bound whilst specification (6) is the upper bound. As these bounds are both negative, this provides additional evidence that there is a negative impact of absences on achievement.

In summary, using a multi-dimensional fixed effect model, this research finds the impact of a session absent on achievement in Maths and English to be negative and significant. These findings support Becker (1962) Human Capital Theory, assuming that each session at school is seen as an investment in education.

The magnitude is small compared to other factors considered in the education production function literature. For example, Krueger (2003) cost-benefit analysis of class size, estimates the impact of a 7 pupil reduction to the average class size would result in a 0.10 standard deviation increase in productivity. Dickerson et al. (2018) estimate that a 1 standard deviation increase in peer quality is associated with a 3-5 percentage point increase in the likelihood

[32] We also run a model which include exclusions to control for behaviour. Within the sample, 87.5% are never excluded. There is also very little within individual variation. We therefore argue that the individual fixed effect is controlling for behaviour.

[33] To check whether the estimates are driven by changes in the sample size, we estimate specifications 1-4 using the sample from specification 5. The results presented in Table B.1 and B.2 in the Appendix show that the estimates are not affected by changes in the sample size.

of pupils aspiring to academic rather than vocational courses.

However, even though absences are analysed across various countries and time periods, the results presented here fall within the range of estimates found in the instructional time literature and are comparable to those inGottfried (2010), Goodman (2014), Aucejo & Romano (2016), Cattan et al. (2023) and Liu et al. (2021).[34]

Aucejo & Romano (2016) do find absences to have a larger impact on Maths than reading in North Carolina. In contrast, in our context, the results presented in Tables 3.3 and 3.4 suggest a slightly larger impact on Maths than on English, though the coefficients are not significantly different from one another.

Table 3.3: The impact of absences on end of year maths achievement

	(1)	(2)	(3)	(4)	(5)	(6)
Absence	-0.0126***	-0.0097***	-0.0063***	-0.0042***	-0.0034***	-0.0065***
	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0009)	(0.0006)
Pupil FE	No	No	No	Yes	Yes	No
School FE	No	No	No	No	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No	Yes
N	9279	9279	5187	9279	6869	4236

Notes: Dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. All specifications include dummy variables for key stage, year, free school meal eligibility and special educational needs. Specifications without individual fixed effects also include dummy variables for race, gender and birth month. Controls include parents' characteristics, including education and net monthly income. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[34]The literature has consistently found that absences have negative effects on test scores. Though there has been large inconsistencies in the magnitude, results have ranged from -0.045% of a standard deviation to -5%.

Table 3.4: The impact of absences on end of year english achievement

	(1)	(2)	(3)	(4)	(5)	(6)
Absence	-0.0113***	-0.0087***	-0.0059***	-0.0042***	-0.0036***	-0.0060***
	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0009)	(0.0007)
Pupil FE	No	No	No	Yes	Yes	No
School FE	No	No	No	No	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No	Yes
N	9215	9215	5135	9215	6784	4188

Notes: Dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. All specifications include dummy variables for key stage, year, free school meal eligibility and special educational needs. Specifications without individual fixed effects also include dummy variables for race, gender and birth month. Controls include parents' characteristics, including education and net monthly income. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5.2 Functional form

The baseline model assumes that pupil absences incur neither diminishing nor increasing costs and that there is no abrupt change in effect at the threshold for being classified as persistently absent, which was set at 20% prior to 2011. The threshold was then lowered to 15% in 2011.[35]

The next stage of the analysis involves examining how pupil absences might have non-linear effects on current achievement. We may expect there to be a non-linear effect as pupils might be able to catch up on a few sessions absent but this may become more difficult the more absences a pupil has. To investigate the presence of a non-linear effect, this research adds a quadratic term to specification (5).

[35] In 2015, the threshold to be classed as persistently absent was lowered again to 10%. This falls outside of this research sample period.

Table 3.5: The non-linear impact of absences on end of year maths and english achievement

	(Maths)	(English)
Absence	-0.0038*	-0.0042**
	(0.0017)	(0.0014)
Absence^2	0.0032	0.0054
	(0.0149)	(0.0079)
N	6869	6784
R^2	0.65	0.53

Notes: Dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. The displayed results are from specification 5 which includes pupil and school fixed effects. The specification also includes dummy variables for key stage, year, free school meal eligibility and special educational needs alongside parents' characteristics, including education and net monthly income. The quadratic term has been divided by 1000. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The coefficient on the quadratic term is positive, small in magnitude, and highly insignificant.

To test whether there is a discontinuity at the threshold of being considered persistently absent, we introduce a dummy variable equal to one if the individual has a 20% absent rate or in other words, 72 or more absent sessions. Table 3.6 provides the estimates which are negative and significant at a 10% significance level. Individuals who are classed as persistently absent have levels of achievement 24% of a standard deviation less for Maths and 38% for English than those with lower levels of absence. These findings are comparable to the estimates for one session absent in Table 3.3/ 3.4 where the estimated impact on one session absent was -0.0034/-0.0036. This provides further evidence that the relationship is linear.

Table 3.6: Discontinuity at persistently absent threshold

	(Maths)	(English)
Absence 72	$-0.2416^{\dagger}$	$-0.3044^{\dagger}$
	(0.1355)	(0.1736)
N	2786	2738
R^2	0.40	0.14

Notes: Dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. The displayed results are from specification 5 which includes pupil and school fixed effects. The specification also includes dummy variables for key stage, year, free school meal eligibility and special educational needs alongside parents' characteristics, including education and net monthly income. Standard errors are clustered at the school level and are shown in parentheses.

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

These findings suggest that the marginal effect of absences is approximately linear in both Maths and English. This finding is in line with Cattan et al. (2023), Aucejo & Romano (2016), and Liu et al. (2021) who also do not find evidence of non-linear impacts.

The data used in this research does not report the duration of absences. Absences could be spaced out over the year or concentrated in one long absence period. It could be expected that absences which are spaced out may be easier to mitigate than long continuous periods of absence. This assumption is untested in this research but could be used to explain the linear findings presented here.

3.5.3 Heterogeneity

On average, school absences have a negative effect on end of year academic attainment in Maths and English. The next stage of this analysis explores the extent to which the effect of pupil absences differs across observable characteristics. Understanding heterogeneous impacts is important for policy development as it can help to identify groups of pupils who may benefit more from intervention.

We extend specification 5 by adding interaction terms to explore how the impact of absences might differ based on various pupil and family characteristics. Column 1 in the tables below reports the estimates for Maths and column 2 does so for English.

We begin by examining how the negative impact of absences varies across different demographic groups. Tables 3.7 and 3.8, estimate the impact of absences across gender and ethnicity.

This is driven by evidence of differences in achievement in Maths and English within schools, based on gender and ethnicity. Borgonovi et al. (2021) provide evidence to suggest that girls outperform boys in English whilst this is reversed for Maths. A government-published report into race and ethnic disparities shows that all ethnic minority groups apart from Black Caribbean perform better than white pupils at school (Commission on Race and Ethnic Disparities 2021).

Table 3.7: The impact of absences on end of year academic achievement, by gender

	(Maths)	(English)
Absence	-0.0032*	-0.0002
	(0.0016)	(0.0018)
Absence*Female	-0.0002	$-0.0022^{\dagger}$
	(0.0010)	(0.0012)
N	6869	6784

Notes: Dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. The displayed results are from specification 5 which includes pupil and school fixed effects. The specification also includes dummy variables for key stage, year, free school meal eligibility and special educational needs alongside parents' characteristics, including education and net monthly income. Standard errors are clustered at the school level and are shown in parentheses.

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

In Table 3.7, the coefficient for absences represents the effect on men, while the interaction term indicates the difference in impact between men and women. The negative interaction term suggests that absences have a greater impact on females. However, both interaction coefficients are small and only significant for English at a 10% significance level.

Table 3.8 presents the coefficients on the ethnicity interactions. The coefficient on absences shows the impact for White individuals and the interactions show the difference in impact between white and Asian, black, mixed, and other ethnic minorities. The sign on the interaction coefficients, apart from the impact on Maths test scores for Asian pupils and pupils with an other ethnicity, is positive. This suggests that absences are less harmful to non-white pupils. Though, as before, the differential estimates are highly insignificant.

Table 3.8: The impact of absences on end of year academic achievement, by ethnicity

	(Maths)	(English)
Absence	-0.0034**	-0.0039***
	(0.0011)	(0.0010)
Absence*Asian	-0.0003	0.0023
	(0.0030)	(0.0033)
Absence*Black	0.0002	0.0016
	(0.0030)	(0.0042)
Absence*Mixed	0.0000	0.0020
	(0.0023)	(0.0023)
Absence*Other	-0.0063	-0.0106
	(0.0079)	(0.0105)
N	6869	6784

Notes: See Table 3.7 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To summarise, this research finds no evidence of heterogeneous impacts of absences across gender or ethnicity. These findings are in line with Cattan et al. (2023) and Liu et al. (2021).

The impact of absences may also differ across household characteristics. Household income may determine how much support the parent is able to give their children. Higher income could mean that parents have more demanding jobs and therefore less free time to support their children. However, they may be able to afford private tuition. Those with lower income may not have the skills to help their children with school work.

Table 3.9 provides the results for absences interacted with income categories based on percentiles. The coefficient on absence indicates the impact for those pupils with family income below the 25th percentile. The interactions show the difference in impact between low income and median income and between low income and income above the 75th percentile.

The impact of absences does seem to decrease as income increases[36]. However, the difference in impact is not statistically significant so we conclude that we find no evidence of heterogeneous impact across parental income.[37] Parents' income does not seem to be a substitute for in person teaching.

Table 3.9: The impact of absences on end of year academic achievement, by parental income

	(Maths)	(English)
Absence	-0.0035***	-0.0037***
	(0.0010)	(0.0009)
Absence*Median	-0.003	0.0003
	(0.0013)	(0.0015)
Absence*75% percentile	0.0031	0.0026
	(0.0023)	(0.0023)
N	6869	6784

Notes: See Table 3.7 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Catching up on missed learning due to an absence is likely to be more difficult for pupils who are of lower ability. We take specification (4)[38] and run a quantile regression. Figure 3.7 shows the impact of absences across the Maths achievement distribution whilst Figure 3.8 is for English. Figure 3.7 shows that for Maths, those at the lower end of the ability distribution are impacted by absences more than those higher up the ability distribution. Whilst for English, there does not seem to be much difference in the impact of absences across the

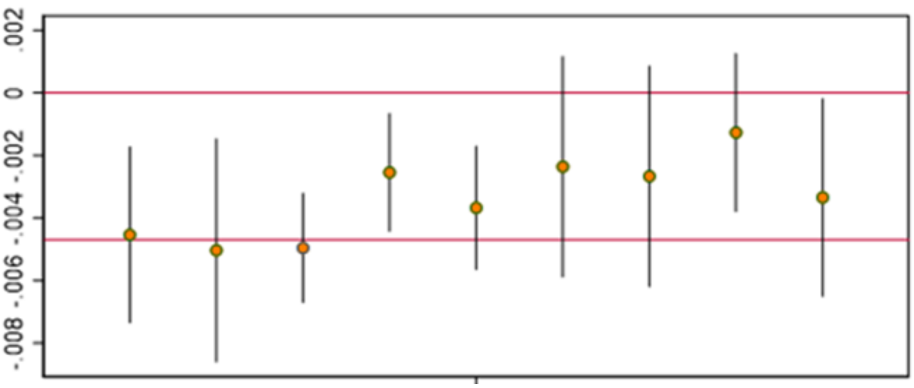
[36]The estimated coefficient on the interaction at the 75th percentile nearly offsets the base effect. However, this is insignificant. In future research it would be useful to investigate this with a larger dataset.

[37]We found identical results using other cuts of the parent's income distribution as well as a continuous by continuous interaction.

[38]Specification (4) is used as opposed to specification (5) as there is currently no estimation strategy for a quantile regression with multiple fixed effects.

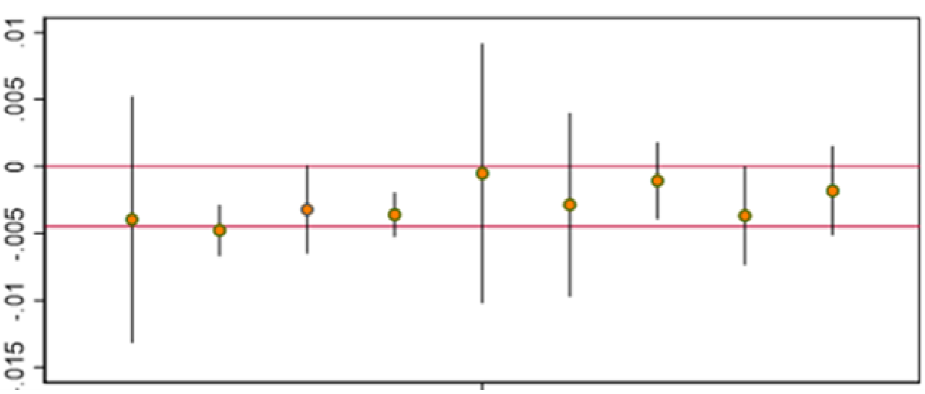
ability distribution. It may be easier to catch up on missed learning in English whereas Maths builds on prior knowledge and so missed learning through absences has more effect. Parents could also be more able to help their children catch up with missed learning in English than Maths.

Figure 3.7: Impact of absences by maths ability



Note: Observations are counts of yearly absent sessions across year 6, year 9 and year 11. The ability variable is separated in to 9 quantiles. The first horizontal line is set at 0, the second is set at the estimate produced in specification 4 in Table 3.3.

Figure 3.8: Impact of absences by English ability



Note: Observations are counts of yearly absent sessions across year 6, year 9 and year 11. The ability variable is separated in to 9 quantiles. The first horizontal line is set at 0, the second is set at the estimate produced in specification 4 in Table 3.4.

Finally, Table 3.10 examines heterogeneous effects by school year. The coefficient on absences shows the impact for pupils in Year 6 and the interactions show the difference in impact between Year 6, Year 9, and Year 11. The interaction coefficients are positive for Maths suggesting that the impact of absences is less in higher year groups. For English, the interaction for year 9 is positive whilst for year 11, it is negative, though, all interaction coefficients are insignificant for this subject. Pupils' capacity to learn may increase with age so the impact of an absence declines as they get older.

Table 3.10: The impact of absences on end of year academic achievement, by school year

	(Maths)	(English)
Absence	-0.0067*	-0.0031
	(0.0032)	(0.0032)
Absence*Y9	0.0071*	0.0016
	(0.0033)	(0.0033)
Absence*Y11	0.0016	-0.0015
	(0.0032)	(0.0033)
N	6869	6784

Notes: See Table 3.7 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Although this section does not reveal evidence of varying impacts across key demographic characteristics, it does show that low ability pupils, especially in Maths, would gain the most from reduced absences. Consequently, these findings suggest that increasing instructional time could help narrow performance gaps.

3.5.4 Robustness Checks

Alternative measures of achievement

As discussed in Section 3.3, NPD records KS3 and KS4 achievement in multiple ways. Prior to 2008, KS3 SATS were taken by all pupils in Year 9. These tests were externally set and marked. The baseline analysis uses National curriculum levels which are teacher assessed. As we are unable to control for teacher characteristics in this research, there could be omitted variables, for example, pupil teacher relationships, which could bias the estimates.

As a robustness check, we estimate the same specifications as set out in Section 3.5.1 but restrict the samples to those who took KS3 SATS. Table 3.11 provides the estimates for Maths whilst Table 3.12 does so for English.

Table 3.11: The impact of absences on end of year maths achievement (using KS3 SATs)

	(1)	(2)	(3)	(4)	(5)	(6)
Absence	-0.0116***	-0.0089***	-0.0070***	-0.0045***	$-0.0045^{\dagger}$	-0.0055***
	(0.0006)	(0.0006)	(0.0009)	(0.0009)	(0.0018)	(0.0016)
Pupil FE	No	No	No	Yes	Yes	No
School FE	No	No	No	No	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No	Yes
N	6634	6634	1442	6634	3044	737

Notes: See Table 3.3 $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

Table 3.12: The impact of absences on end of year english achievement (using KS3 SATs)

	(1)	(2)	(3)	(4)	(5)	(6)
Absence	-0.0110***	-0.0084***	-0.0058***	-0.0053***	-0.0052***	-0.0061***
	(0.0007)	(0.0007)	(0.0009)	(0.0008)	(0.0014)	(0.0014)
Pupil FE	No	No	No	Yes	Yes	No
School FE	No	No	No	No	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No	Yes
N	6575	6575	1415	6575	2986	709

Notes: See Table 3.3 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Despite the change in the measure of achievement and consequently, the change in the sample analysed, the estimates remain similar.

In the baseline analyses we focus on Maths and English separately. We use the grade achieved in Maths and English in GCSE exams as the measure of achievement for Year 11. NPD also records the total point score achieved in all GCSEs taken.

We now estimate the same specifications as set out in Section 3.5.1 but we focus on analysing the impact of absences on total achievement measured at the end of the year. Achievement in KS2 is measured as the total point score in Maths and English. For KS4, it is the total point score achieved for all GCSEs taken. In Table 3.13, KS3 achievement is measured using teacher assessments in Maths and English whilst in Table 3.14, the sample is restricted to those who took KS3 SATS.

The estimated impact of absences on achievement is negative and significant. The results show comparable patterns across tables.

Table 3.13: The impact of absences on end of year academic achievement (using KS3 teacher assessments)

	(1)	(2)	(3)	(4)	(5)	(6)
Absence	-0.0129***	-0.0106***	-0.0082***	-0.0059***	-0.0056***	-0.0079***
	(0.0005)	(0.0005)	(0.0004)	(0.0006)	(0.0010)	(0.0006)
Pupil FE	No	No	No	Yes	Yes	No
School FE	No	No	No	No	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No	Yes
N	9279	9279	5202	9279	6869	4252

*Notes: See Table 3.3** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.14: The impact of absences on end of year academic achievement (using KS3 SATs)

	(1)	(2)	(3)	(4)	(5)	(6)
Absence	-0.0129***	-0.0107***	-0.0090***	-0.0071***	-0.0059***	-0.0078***
	(0.0005)	(0.0005)	(0.0007)	(0.0008)	(0.0016)	(0.0014)
Pupil FE	No	No	No	Yes	Yes	No
School FE	No	No	No	No	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No	Yes
N	6649	6649	1457	6649	3072	743

*Notes: See Table 3.3** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The evidence presented here indicates that the baseline results are robust to alternative measures of the dependent variables. The results also provide some evidence to suggest the teacher marking of KS3 assessments is not biasing the estimates.

Timing of absences

As discussed in Section 3.4, one threat to establishing causality is measurement error. We observe the total number of absences in a year, but we do not know the exact timing of those absences. Hence, some of the absences recorded could have occurred after the end of year tests. This over recording of the number of absences could lead to an underestimation of the impact of absences on contemporaneous achievement.

NPD does record the term that the absence occurred. Since most exams take place towards the beginning of the summer term, assuming that the majority of the summer term absences take place after the exams, they should not impact achievement in the same year. If absences occurring after the exam affect test scores, it would indicate that the identification strategy may not be effectively accounting for selection into absences. If post-exam absences have a significant impact on test scores, it suggests that the issue might not be the absences themselves but rather an omitted confounding factor, such as disengagement, that could be driving the negative relationship between absences and exam scores.

Table 3.15 re-estimates specification 5, disaggregating absences in to autumn, spring, and summer absences. For both Maths and English, absences across the year have a negative impact though only absences during the spring term are significantly related to outcomes. These findings are comparable to those by Gottfried & Kirksey (2017) who find that only spring absences have a significant association with test scores. As their data provide the day the absences occurred, they are also able to show that the closer to the exam the absence occurs the greater the adverse impact.

The lack of significance on the relationship between summer absences and achievement provides confidence that the act of missing school itself could lead to lower test scores.[39] The lack of significance on the relationship between autumn absences and achievement suggests that absences that occur earlier in the year can be made up, whilst absences that occur close to the exam are detrimental to achievement.

Table 3.15: The impact of absence of end of year academic achievement, by term

	(Maths)	(English)
Autumn	-0.0017	-0.0020
	(0.0016)	(0.0017)
Spring	-0.0043*	-0.0042*
	(0.0018)	(0.0018)
Summer	$-0.0051^{\dagger}$	-0.0053
	(0.0028)	(0.0034)
N	6869	6784
R^2	0.64	0.62

Notes: See Table 3.7 $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

Time-varying unobservables

The main estimation technique used in this research is fixed effect analysis. The causal interpretation of the estimates is pinned on the assumption that there are no time-varying unobservables that impact both absences and attainment.

As absences are largely driven by illness, a particular threat to the identification assumption would be an idiosyncratic health shock. To test for this, we estimate specification (5) but split absences into authorised and unauthorised. If a health shock were influencing the results, we would expect that, after separating absences into authorised and unauthorised types, authorised absences would have a more negative effect compared to unauthorised absences.

[39]For Maths, summer absences are significant at a 10% significance level. Summer absences will include absences very close to the exams as well as after the exam period.

Table 3.16: The impact of absence of end of year academic achievement, by type

	(Maths)	(English)
Authorised	-0.0026**	$-0.0019^{\dagger}$
	(0.0009)	(0.0011)
Unauthorised	-0.0048**	-0.0062***
	(0.0017)	(0.0016)
N	6869	6784

Notes: See Table 3.7 $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

The results presented in Table 3.16 suggest that an additional authorised absence is associated with a fall in Maths (English) achievement by 0.26% (0.19%) of a standard deviation while unauthorised absences have an impact of 0.48% (0.62%) of a standard deviation.[40] These results indicate that health shocks do not seem to bias the results.

Reverse causality

Another threat to establishing causality discussed in Section 3.4, is reverse causality. Pupils who gain low achievement scores could be put off from attending school. If this is the case, then the estimates presented in Table 3.3 and 3.4 will not be causal.

To test this possibility, we estimate the impact of absences in the previous year on subsequent achievement the argument being that previous absences are predetermined and so not a function of current achievement.

The results presented in Table 3.17 suggest a likely persistent effect, whereby the previous year's absences impact subsequent years' academic achievement even after controlling for lagged test scores. This provides evidence against the reverse causality argument.

[40]The impacts of authorised and unauthorised absences on achievement in Maths are not statistically different from each other.

Table 3.17: The impact of absence of end of year academic achievement including the lag of absences

	(Maths)	(English)
Absence_{t-1}	-0.0049***	-0.0025**
	(0.0010)	(0.0010)
N	4252	4222
R^2	0.64	0.54

Notes: See Table 3.7 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sample selection

As discussed in Section 3.3, parents are asked for consent to link their children's education records in NPD to survey data in UKHLS. We may therefore have concerns regarding sample selection bias. Parents of children who are of high ability may be more likely to consent to the linkage. Parents who are more aware of the benefits of the linkage may also be more likely to consent. If children of higher ability also have lower absences, this would create a negative bias.

Al Baghal (2016) analyses factors influencing consent outcomes in UKHLS. They provide evidence to show that nearly all mothers give the same consent outcomes for all their children. This suggests that child-level factors may not be important factors in determining consent. Parental characteristics, including race, socioeconomic status, political party support and trust were found to be the most influential factors. Al Baghal (2016) finds that minority ethnicities and parents who have a degree were less likely to give consent, whilst those supporting left-wing political parties and expressing a greater propensity to trust strangers were more likely to give consent. This research controls for race and parents' education level which Al Baghal (2016) found to be the largest determining factors of consent. We also argue that political party support and trust are unlikely to be large determinants of children's education outcomes. Table 3.1 also shows that the sample used in this research is representative of the English population.

The evidence discussed in Section 3.5.4 shows that the baseline results remain consistent across various specifications, indicating that potential threats to the identification strategy are unlikely to be influencing the results.

Oster bounds

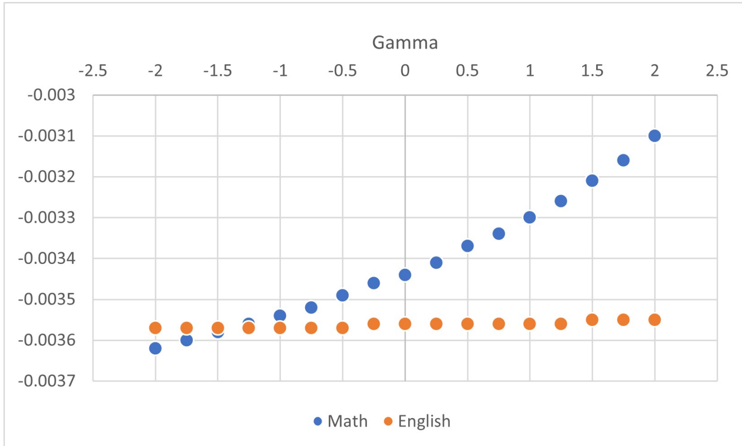
In Section 3.5.4 we argued that the main threat to the identification assumption would be an idiosyncratic health shock. While this may be the most common threat, there is still potential for other time-varying individual factors to bias our estimates. We therefore also provided bounds, using Oster (2019) method[41], to allow for extreme correlation between absences and the unobservables that we cannot control for.

Since selection based on unobservables could go in the same or opposite direction as selection on observables, we account for the possibility that the true effect might be either overestimated or underestimated. We estimate the bounds assuming gamma takes a value of 2 to -2. This allows use to use the extreme assumption that selection of the unobservables is twice as strong as selection of the observables.

We can see from Figure 3.9 that estimated bounds for the effect of absences on academic performance are -0.31% to -0.36% for Maths and remains fairly constant for English. Therefore, even with large levels of unobserved selections, the effect of half a day absence is very close to our baseline estimate and remains fairly robust despite the level of unobservable imposed.

[41] The Oster (2019) method is a statistical technique designed to obtain beta bounds for a coefficient of interest. It works by combining coefficient stability with information about R-squared movements to estimate how much bias might come from the missing variables.

Figure 3.9: The impact of absences on achievement at different values of gamma.



Note: The displayed results are from Specification 5 which includes individual and school fixed effects. We assume an $R^2 = 0.6$.

3.6 Conclusion

Using novel data from England and a multi-dimensional fixed effect model, we provide the first study to shed light on the causal impact of absences on contemporaneous achievement, in the context of England.

Our analysis shows that on average missing one additional session of school reduces an individual's end of year attainment in Maths/English by 0.3% of a standard deviation. Our findings deliver very robust results across different samples of individuals and multiple measures of achievement. The robustness checks consistently suggest that the existence of an omitted variable bias, reverse causality or measurement error are unlikely which supports a causal interpretation. The findings are also comparable to the literature.

This research has important policy implications. Firstly, consistent with the literature, we find approximately linear effects of pupil absences on end of year achievement in Maths and English. We also find no evidence of a discontinuity at the persistently absent threshold.

This linearity suggests that the persistently absent indicator is arbitrary and may miss a large proportion of absences that are having a negative impact on achievement.

Secondly, the UK government, in the 2022 School's White Paper (Department for Education 2022C), announced the introduction of a minimum expectation on the length of the school week of 32.5 hours. It is not currently possible to analyse the impact of extending the school week in England as The Department for Education does not record a schools' opening hours. However, the ability to target specific pupils who would benefit the most from reduced absences suggests that focused initiative might be more effective than extending the school week for everyone. Additionally, this could also help narrow current achievement gaps.

We provide evidence suggesting that reducing absences among pupils at the lower end of the ability distribution could help narrow performance gaps. Our findings show that lower ability pupils miss 11 more days than their higher ability peers. Additionally, we observe varying impacts across the ability distribution, particularly in Maths. Reducing absences for lower ability pupils, to the level of their higher ability peers, could improve the average achievement of low achievers by 3.3% of a standard deviation. While this alone may not fully close the performance gaps, these reductions are substantial.

The significant policy focus on school absences has been largely influenced by the ongoing effects of the COVID-19 pandemic. Despite the shift to online learning, numerous uncertainties, delays, and variations in teaching quality, school closures resulted in substantial learning losses. According to Ofqual (2021), by autumn 2021, primary school pupils had experienced a loss of 1.9 months in Maths and 0.8 months in reading. Disadvantaged pupils experienced an additional loss of 0.3 months in Maths and 0.4 months in reading. Back of the envelope calculations using baseline estimates, assuming linearity, suggest that COVID-related school disruptions could have led to a 24% reduction in attainment in Maths and English, measured as a standard deviation. These findings may only be partially relevant to the COVID situation as we focus on the impact of individual absences, as opposed to school closures. However, as countries started to open up, pupil absence became increasingly driven by individual pupils self-isolating, therefore the true impact of COVID could be much larger than estimated here and could have long-term economic consequences if not appropriately compensated.

In future research, it is worthwhile to explore why pupils are absent and examine interventions that aim to improve pupils' school attendance. Education Endowment Foundation (2022) published a rapid evidence assessment examining the existing research and found that the overall quality of evidence is weak, and mainly from the US.

This research has been able to quantify the impact of absences and show that even low levels of absence could have a larger impact on individual outcomes than we might have expected.

Appendix B

Appendix

Table B.1 provides a summary of the population eligible to receive the consent question at Waves 1 and 4, categorised by whether consent was for a child or an adult, and by gender, along with the number and proportion of respondents who agreed to education data linkage. Consent rates were higher for adults than for children at both Waves 1 and 4. The number of adults for whom data is available in the NPD grew between Waves as more children reached that age range.

Figure B.1: Population eligible to receive the consent question in UKHLS

	Children aged 4-16			Adults 16+		
	All	girls	boys	all	women	men
Wave 1						
Eligible population	9,745	4,743	5,002	5,041	2,758	2,283
Consented	6,480	3,131	3,349	3,915	2,141	1,774
% of eligible	66.5	66.0	67.0	77.7	77.6	77.7
Matched	5,331	2,608	2,723	2,193	1,174	1,019
% of consent	82.3	83.3	81.3	56.0	54.8	57.4
% of eligible	54.7	55.0	54.4	43.5	42.6	44.6
Wave 4						
Eligible population	7,617	3,697	3,920	6,664	3,599	3,065
Consented	4,739	2,344	2,395	5,260	2,892	2,368
% of eligible	62.2	63.4	61.1	78.9	80.4	77.3
Matched	4,678	2,311	2,367	4,700	2,545	2,155
% of consent	98.7	98.6	98.8	89.4	88.0	91.0
% of eligible	61.4	62.5	60.4	70.5	70.7	70.3

B.0.1 Distribution of absent sessions by Key Stage

Figures B.2, B.3 and B.4 show the distribution of (authorised) absent sessions in each school year analysed. The average number of sessions missed in year 6 is 13.6 which is equivalent to 7 days. The median is 10 sessions. 6.89% of pupils in year 6 have no absences and 51.74% miss 10 days or less. The average number of sessions missed in year 9 is 18.6 which is equivalent to 9.5 days. The median is 13 sessions. 6.19% of pupils in year 9 have no absences and 41.66% miss 10 days or less. The distribution of absences for year 9 is wider than year 6. Fewer pupils have no absences, and more pupils have the number of absent sessions exceeding 60. This pattern continues in to Year 11 where the average number of sessions missed in year 11 is 19.5 which is equivalent to 9.75 days. The median is 14 sessions. 5.93% of pupils in year 9 have no absences and 40.69% miss 10 days or less. The number of absent sessions is increasing by year group, and this is also leading to more unauthorised sessions.

Figure B.2: Distribution of (authorised) absent sessions in Year 6

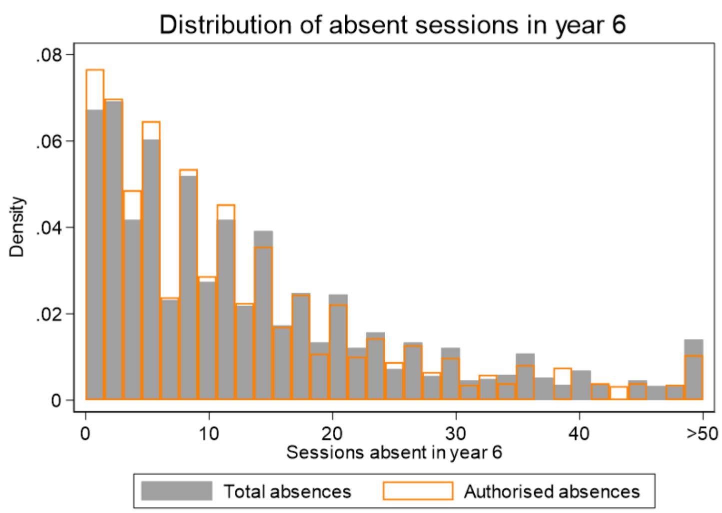


Figure B.3: Distribution of (authorised) absent sessions in Year 9

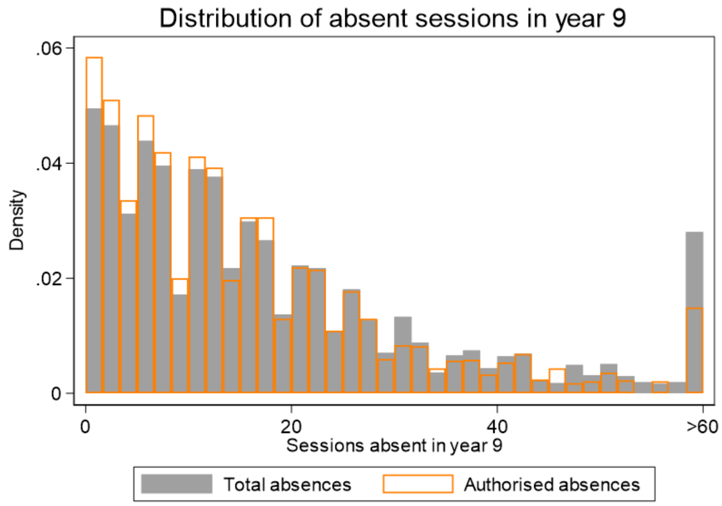
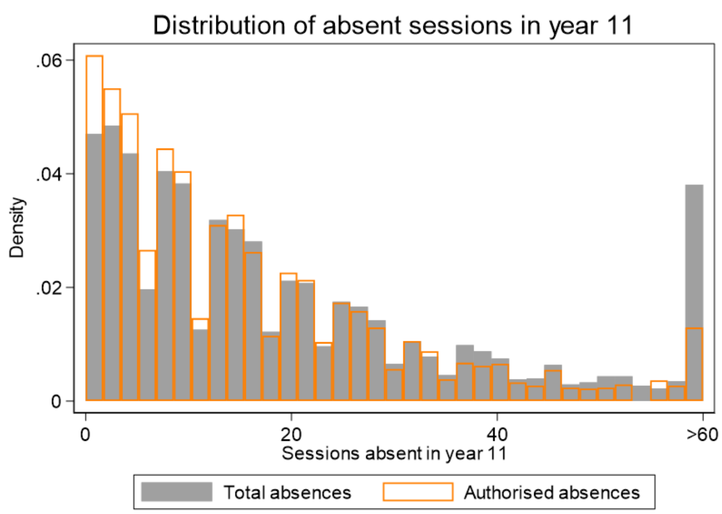


Figure B.4: Distribution of (authorised) absent sessions in Year 11



B.0.2 Correlations by Key Stage

Figure B.5, B.6 and B.7 show the relationship between absences and achievement by Key Stage. We can see that the relationship is negative for all key stages. Though, the correlation is slightly more negative for KS2 and KS3 than KS4 which could be explained as pupils are more independent in KS4 and are able to catch up with school work they missed more easily.

Figure B.5: The descriptive relationship between absences and academic performance in KS2

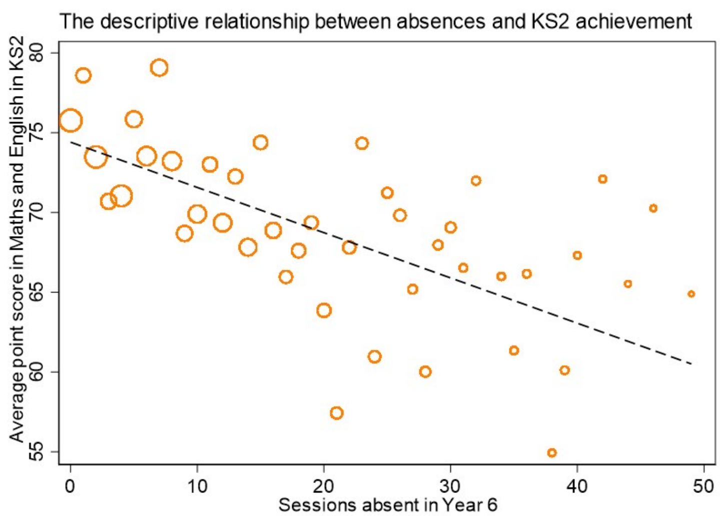


Figure B.6: The descriptive relationship between absences and academic performance in KS3

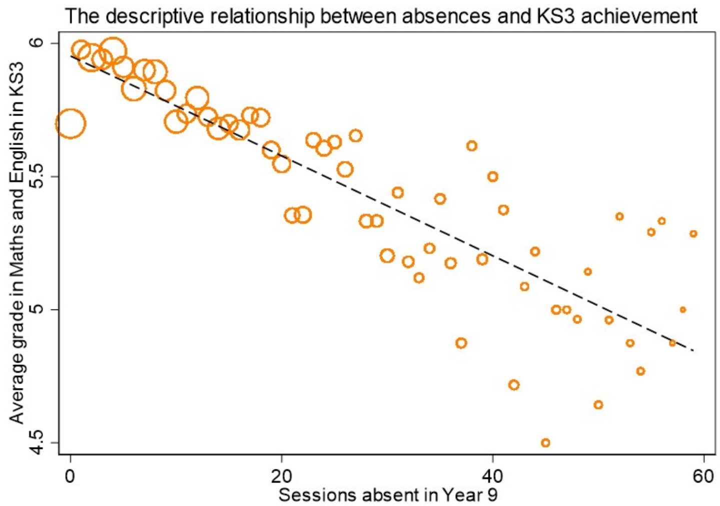
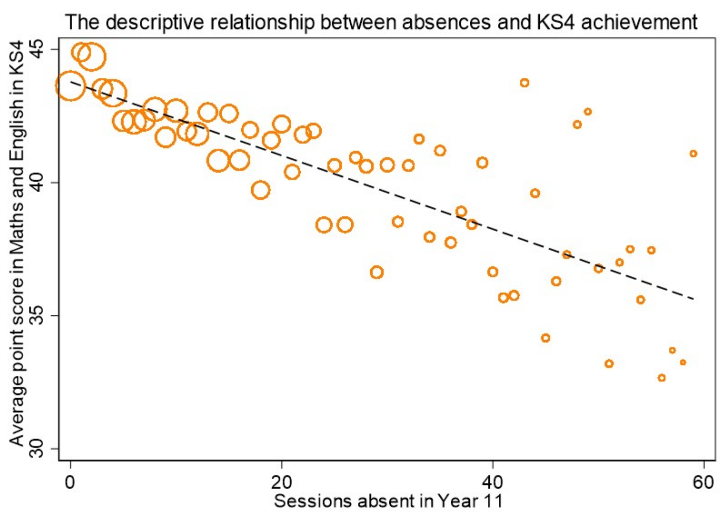


Figure B.7: The descriptive relationship between absences and academic performance in KS4



B.0.3 Test score distributions by subject and Key Stage

Figure B.8: Distribution of maths test scores in KS2

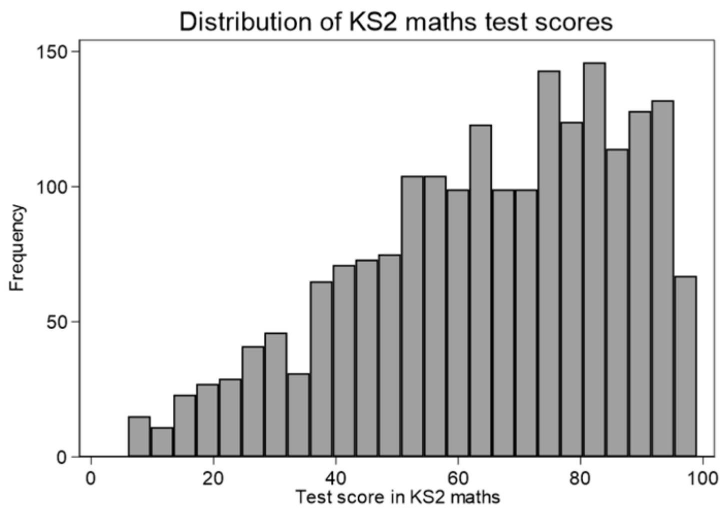


Figure B.9: Distribution of English test scores in KS2

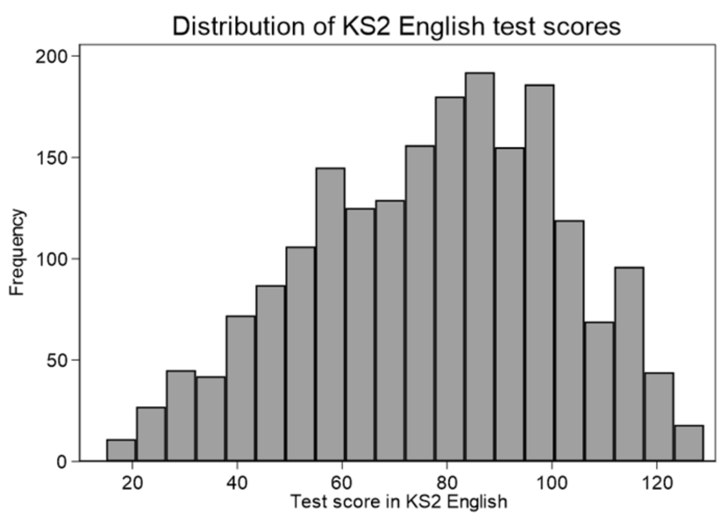


Figure B.10: Distribution of maths test scores in KS3

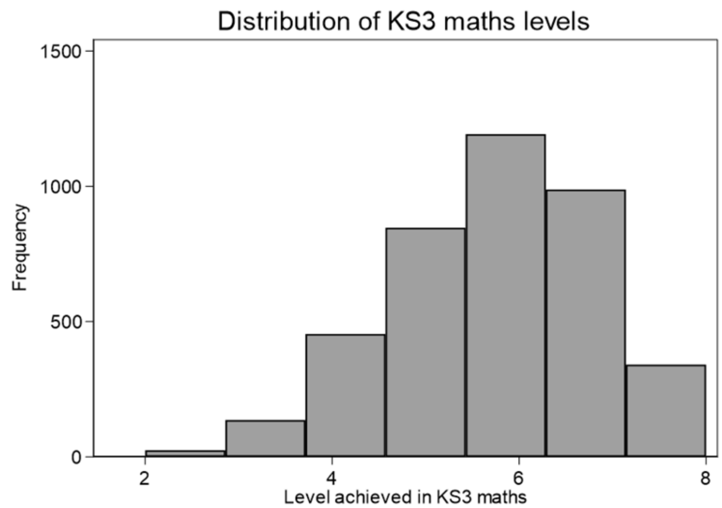


Figure B.11: Distribution of English test scores in KS3

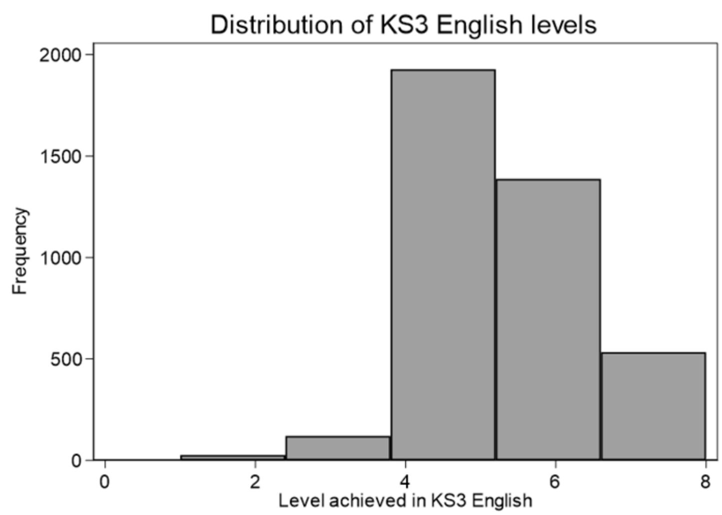


Figure B.12: Distribution of maths test scores in KS4

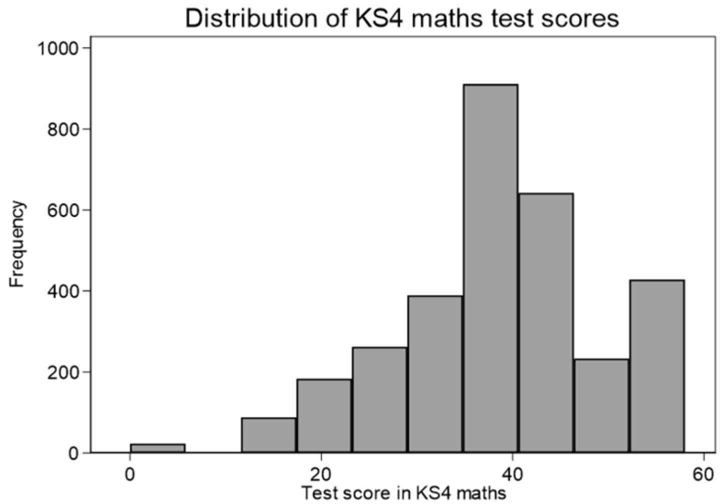
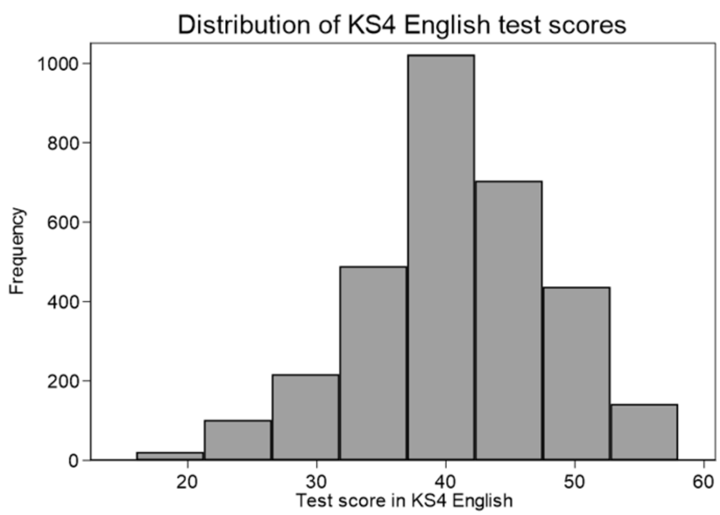


Figure B.13: Distribution of English test scores in KS4



B.0.4 Short run results using the sample from specification 5 in Table 3.3 and 3.4

Table B.1: The impact of absence of end of year maths achievement, focusing on smaller sample

	(1)	(2)	(3)	(4)	(5)
Absence	-0.0124***	-0.0097***	-0.0064***	-0.0037***	-0.0034***
	(0.0006)	(0.0006)	(0.0005)	(0.0007)	(0.0009)
Pupil FE	No	No	No	Yes	Yes
School FE	No	No	No	No	Yes
Controls	No	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No
N	6869	6869	4165	6869	6869

Notes:The table presents the estimates of replicas of Table 3.3 using the sample for specification 5. The dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. All specifications include dummy variables for key stage, year, free school meal eligibility and special educational needs. Specifications without individual fixed effects also include dummy variables for race, gender and birth month. Controls include parent's characteristics, including education and net monthly income. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.2: The impact of absence of end of year English achievement, focusing on smaller sample

	(1)	(2)	(3)	(4)	(5)
Absence	-0.0112***	-0.0089***	-0.0057***	-0.0039***	-0.0036***
	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0009)
Pupil FE	No	No	No	Yes	Yes
School FE	No	No	No	No	Yes
Controls	No	Yes	Yes	Yes	Yes
Lagged achievement	No	No	Yes	No	No
N	6712	6712	4070	6712	6712

Notes: The table presents the estimates of replicas of Table 3.4 using the sample for specification 5. The dependent variable is standardised by key stage with a mean of 0 and a standard deviation of 1. All specifications include dummy variables for key stage, year, free school meal eligibility and special educational needs. Specifications without individual fixed effects also include dummy variables for race, gender and birth month. Controls include parent's characteristics, including education and net monthly income. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 4

Exploring Diversity: The impact of ethnic minority peers on the perceived likelihood of attending university for white pupils

4.1 Introduction

Many places in Britain are now being described as highly ethnically diverse and residential separation between people of different ethnic groups is decreasing (Catney et al. 2023). Social scientists have long been interested in the impacts of diversity on individual and societal outcomes. One of the significant differences between ethnic groups is university participation. On average, all ethnic minority groups are significantly more likely to attend university compared to their White British counterpart. Burgess (2014) was one of the first to attribute these differences to aspirations. The aim of this paper is to investigate the influence of ethnic minorities on white pupils' perceived likelihood of attending university.

Crawford & Greaves (2015) showed that the ethnic minorities have higher university participation rates than White British individuals by at least 5 percentage points, with some groups showing much larger differences. For example, Chinese pupils are, on average, nearly 40 percentage points more likely to attend university compared to White British pupils.[1] This may be surprising considering ethnic minorities in the UK tend to have more disadvantaged backgrounds than White British individuals[2] (a factor typically associated with lower rates of higher education participation). After adjusting for background characteristics and prior attainment, ethnic minority pupils are, on average, much more likely to attend university than their White British counterparts[3]

The large differences in higher education participation rates among different ethnic groups, even after accounting for various background characteristics and prior attainment, suggest that certain factors more prevalent in ethnic minority families, but less common in White British families, are positively linked to higher education participation. Fernández-Reino (2016) and Burgess (2014) have argued that these unobserved factors could be higher aspirations.

Choices regarding higher education can have long-term effects on both economic factors, such as income and job opportunities, and non-economic aspects, including health. There is a large amount of literature on the decision to participate in higher education (Chowdry et al. 2013, Younger et al. 2019). A much smaller literature has examined the influence of peers on this decision.

[1]Approximately 28% of White British pupils attend university at age 18 or 19, compared to around 66% of Chinese pupils.

[2]Over 60% of Bangladeshi individuals fall in to the lowest SES band (bottom 20%), as do around 45% of black African and black Caribbean individuals (Dearden & Sibieta 2010)

[3]Black African pupils are nearly 35 percentage points more likely to attend university than similar White British pupils; most other ethnic minority groups are around 15-25 percentage points more likely to attend university than similar White British pupils (Crawford & Greaves 2015).

Previous literature has provided international evidence on the heterogeneity in educational peer effects. The majority of the literature has focused on the impact of migrants or refugee children on native educational outcomes (Figlio et al. 2024, Figlio & Özek 2019, Morales 2022, Chareyron et al. 2021) However, despite the use of similar methodologies, their findings have varied from positive to negative. Turning to the impact of peers' ethnicity, the causal evidence is limited and mainly focused on the US, again with varying findings. Furthermore, the research on educational peer effects has focused almost exclusively on educational attainment. Of the few studies that do examine the impact on other pupil outcomes, none examine the impact of migrant, refugee status, or ethnicity. As findings vary depending on the country, time frame, and outcome measure used, it is difficult to determine the generalisability of the findings.

The objective of this research is to explore the relationship between peers' ethnicity and individual reported likelihood of attending university for white British pupils, a clear gap in the literature which requires investigation. Throughout this paper we use the term ethnic minorities (not including white minorities) to refer to individuals from the Asian, black, mixed or other ethnic backgrounds.

Against this backdrop we make three contributions. Firstly, we provide the first evidence on the impact of peers' ethnicity on individuals' perceived likelihood of attending university. Individuals' beliefs about further goals play a crucial role in educational and other economic decision making. Studying these beliefs can provide valuable information for policymakers on how to promote economic growth. Analysing the impact of peers' ethnicity sheds light on social dynamics and inequalities. Gorard et al. (2012), who provide a literature review on the relationship between aspirations and educational attainment and participation, describe evidence of a positive association between aspirations and university participation. Secondly, we focus on the impact of ethnic minorities. The literature has mainly examined the impact of refugees or immigrants with native language challenges. Between 2001 to 2021, the percentage of people in England and Wales that identify as white British went down from 87.5% to 74.4% (Office of National Statistics 2024). As the UK becomes more ethnically diverse, it is more important than ever to understand the impact of ethnic diversity on the UK economy. Finally, we create a unique instrument. We argue that the geographical distribution of ethnic minorities today is correlated with job shortages post World War 2. These job shortages, specifically for nurses, were filled by government backed recruitment from Commonwealth countries. We use census data from 1951 to identify areas within the country that had the highest nurse shortage and use this to instrument the proportion of ethnic minorities within a school today.

Identifying causal peer effects is a challenge. As noted by Manski (1993), individuals in the same peer group often display similar behaviours for three primary reasons: first, 'endogenous effects,' where an individual's behaviour is influenced by the group's behaviour; second, 'exogenous effects,' where an individual's behaviour is shaped by the group's external characteristics; and third, 'correlated effects,' where individuals behave alike due to shared traits or similar environments. Additionally, there is the 'reflection effect,' which suggests that any observed correlation between an individual's outcomes and those of their peers might result from the individual influencing their peers, rather than the reverse.

As ethnic minorities are not equally spread across the country, there is a high degree of self-selection into areas. Parents with similar characteristics tend to select into the same areas. This self-selection into peer groups creates selection bias from the fact that an outcome we attribute to a peer effect is just a consequence of the fact that people who share similar characteristics make themselves into groups. Using the proportion of nurses in 1951 as an instrumental variable allows us to overcome this bias as long as the number of nurses in 1951 is correlated with the number of ethnic minorities in an area in 2014 and uncorrelated with the error term. This approach allows us to isolate the variation in the proportion of ethnic minorities in the school that is independent of the error term, thus providing a more accurate estimate of the causal effect.

Utilising the linkage between the Millennium Cohort Study (MCS) and the National Pupil Database (NPD) and exploiting the historical data on occupations, our results can be summarised as follows: the proportion of ethnic minority peers has a positive and significant impact on white pupils' perceived likelihood, at age 14, of attending university. We find these effects are significant for both male and female pupils and larger for the most disadvantaged pupils. The findings are robust to a number of specification tests including weak instrument robust testing, changes in sample size and potential omitted variables.

As the UK government, along with many other countries including France, Italy and the US, promises to cut net migration, these results suggest that through successful integration, diverse populations can have benefits on society.

This research is organised as follows. Section 4.2 reviews the existing literature. Sections 4.3 and 4.4 outline the data and methodology respectively. Section 4.5 presents the estimated impact and tests the robustness, before concluding in Section 4.6.

4.2 Literature Review

4.2.1 Overview

The influence of peers has been analysed across a range of outcomes including; smoking and drug use (Gaviria & Raphael 2001, Kooreman 2007, Argys & Rees 2008), teenage pregnancy (Jackson 2021), obesity (Carrell et al. 2011), take up of government programmes (Rege et al. 2012, Dahl et al. 2014) and financial decisions (Bursztyn et al. 2014). Economists are particularly interested in peer effects as peer group behaviour or characteristics can have an impact on an individual's social and economic decisions that will affect their life chances, including education outcomes (see Paloyo (2020) for a recent literature review) and labour market decisions (Mas & Moretti 2009, Lépine & Estevan 2021).

The economic research on educational peer effects has predominantly focused on the relationship between individuals' and peers' educational attainment (Paloyo 2020). Smaller strands of the literature have analysed the impact of peer characteristics Hoxby (2000) as well as other outcome variables including further education decisions and pupil aspirations (De Giorgi et al. 2009, Dickerson et al. 2018, Gagete-Miranda 2022).

This research aims to combine these three strands of the peer effects literature to examine the impact of ethnic diversity on individuals' perceived likelihood of attending university.

4.2.2 Identification issues

The traditional approach for assessing peer effects involves using observational data to regress individual outcomes on the average outcomes of their peers (Summers & Wolfe 1977).

As detailed in Manski (1993), identifying a causal relationship is a challenge for three main reasons. Firstly, peer groups do not form randomly. Homophily influences the formation of friendships, meaning individuals are more likely to interact with others who share similar traits, such as ethnicity or socioeconomic status (SES) (Currarini et al. 2009, Bramoullé et al. 2012, Currarini et al. 2016, Gagete-Miranda 2022). Secondly, there is a reflection problem. Peers might impact each other simultaneously; this makes it difficult to disentangle if an individual's outcome is the cause or the consequence of peers' behaviour. Finally, peers can be exposed to common inputs, affecting each of their outcomes, such as teachers or school resources which can be difficult to control for, impeding causal interpretation.

Non-random selection into schools is the main identification challenge this literature faces. Parents with similar characteristics tend to select into the same areas. Dustmann & Preston

(2001), using data from England, find that individuals with stronger hostility toward ethnic minorities are less likely to settle in neighborhoods with a high concentration of ethnic minorities. Betts & Fairlie (2001) and more recently, Cascio & Lewis (2012) provide evidence in support of 'white flight', where white pupils move to private schools as a response to increasing levels of ethnic minorities.

However, identifying whether the expected bias on the correlation between exposure to ethnic minorities and peer outcomes is positive or negative is a further challenge. Ethnic minorities, specifically immigrants, are more likely to settle in areas with a lower than average level of education and face native language limitations (Dustmann & Preston 2001). This would lead to a negative bias if lower education levels and language barriers negatively affect individual outcomes. However, ethnic minority parents tend to have high expectations for their children and therefore may select higher quality schools (d'Este & Einiö 2021). Additionally, if parents think that exposure to low SES students and ethnic minorities negatively impacts their children's academic performance, they might choose to send their lower achieving child to a school with fewer non-native students. This would lead to a positive bias if high parental expectations and higher quality schools leads to improved individual outcomes.

4.2.3 Educational peer effects

We begin this literature review by discussing the traditional research on peer effects within education, which analyses the relationship between an individual's and peer's academic performance.

This strand of the literature has analysed data from numerous countries including the US, UK, China, Kenya, and Israel, alongside a host of European countries, using a range of estimation techniques. There is a large amount of literature that exploits exogenous variation in peer quality to identify the effects of peers on individual achievement (Sacerdote 2001, Zimmerman 2003, Booij et al. 2017, Feld & Zölitz 2017, Xu et al. 2021). Others use quasi-experimental research designs such as instrumental variables (Speckesser & Hedges 2017, Mendolia et al. 2018) and regression discontinuity (Ding & Lehrer 2007). Alternatively, researchers have acquired rich datasets (Hoxby 2000, Betts & Zau 2004, Lavy et al. 2007, Burke & Sass 2013) that provide the ability to control for a range of fixed effects alongside a large number of control variables.

Despite the substantial amount of published research that estimates peer effects in education, the literature has been unable to reach a consensus. Sacerdote (2011) and more recently Paloyo (2020) provide in-depth literature reviews on educational peer effects. Sacerdote (2011) notes in his review "results from the myriad peer effects studies would seem to be all over the map." Paloyo (2020) relates this to the highly context-specific nature of the analyses.

Studies have generally found that sharing a classroom with high-achieving peers has positive effects on educational performance (Booij et al. 2017, Feld & Zölitz 2017, Speckesser & Hedges 2017, Mendolia et al. 2018, Hoxby 2000, Burke & Sass 2013, Sojourner 2013), while sharing a classroom with low-achieving peers tends to have negative effects (Figlio & Kenny 2007, Carrell & Hoekstra 2010).

However, the magnitude of these findings varies considerably between studies, subjects, and levels of education. There is a general consensus that peer effects among university roommates are present and of modest size (Sacerdote 2001). However, there is no agreement on the extent of peer effects in primary and secondary education. Feld & Zölitz (2017) and Booij et al. (2017) report that a 1 standard deviation increase in peers' ability leads to increases in individual test scores of 1.26% and 0.95 of a standard deviation, respectively. In contrast, Zimmerman (2003) and Gibbons & Telhaj (2008) find small and insignificant peer effects.

4.2.4 Diversity in peer effects

This research focuses on a sub-section of the peer effects literature which examines the impact of peer characteristics, specifically ethnicity and immigration status.

There is a range of hypotheses to explain the relationship between diversity and pupil outcomes. One, which is used to explain the negative relationship, is decreased teaching efficiency as it can be easier for teachers to deal with a homogeneous group of pupils.

On the other hand, diversity can stimulate the curiosity of pupils as individuals have different skills and interests. Also, an influx of low-skilled migrants may intensify competition in low-qualification sectors, thereby creating additional incentives for natives to enhance their educational outcomes. For the minority group, diversity could boost language assimilation.

4.2.5 Ethnic minorities and non-native peers

Most of the literature examines racial school segregation in the US context (Angrist & Lang 2004, Bifulco et al. 2011, Card & Rothstein 2007, Fryer Jr & Torelli 2010, Hanushek et al. 2009, d'Este & Einiö 2021) or the effects of increased exposure to migrants or refugees in classrooms (Figlio & Özek 2019, Morales 2022, Neymotin 2009, Green & Iversen 2022). In contrast, studies in Europe examining the effects of school segregation are relatively scarce and often concentrate on the concentration of migrant children rather than ethnicity, resulting in mixed findings (Chareyron et al. 2021, Tumen 2021, Ballatore et al. 2018, Bossavie 2020, Hermansen & Birkelund 2015, Geay et al. 2013, Jensen & Rasmussen 2011, Schneeweis 2015).

The literature mainly focuses on secondary school pupils aged 15-16, though, there are a few studies, for example, Geay et al. (2013), Maestri (2017), and Contreras & Gallardo (2022) who focus on younger aged pupils 11–12 year old.

Studies using fixed effects

Fixed effects may control for endogeneity, specifically omitted variables bias, in peer groups as it can be argued that many factors which impact friendship formation are time-invariant, for example, gender, ethnicity, parent education, and socio-economic background.

Hanushek et al. (2009), Maestri (2017), Diette & Uwaifo Oyelere (2017), and Morales (2022) all use variation in the ethnic/non-native composition of adjacent cohorts to estimate the peer effect, controlling for multiple fixed effects including, school, grade, year, and second-level interactions among these three components. The remaining variation in racial composition is due to students switching schools and ongoing differences between cohorts that reflect random demographic changes in the makeup of each cohort within the schools.

Despite using similar methodologies, their findings vary.

Hanushek et al. (2009) estimates show that a 10-percentage point increase in the percentage of Black peers would reduce annual achievement growth for Blacks by 0.025 standard deviations, whilst for whites and Hispanics, the finding is small and insignificant. The share of Hispanic enrollment has a positive but statistically insignificant effect on the achievement of Black and White students, suggesting that it is the proportion of Black students, rather than the proportion of ethnic minorities overall, that negatively affects achievement.

Maestri (2017) supports these findings by estimating a negative and statistically significant impact of the share of immigrants on native pupils' Maths and Reading scores, whilst ethnic diversity positively impacts native pupils' language and Maths test scores. These findings are, however, statistically insignificant.

Diette & Oyelere (2014) suggest that the negative impact of limited English (LE) pupils may be focused on males. They find a small negative effect on Black males' scores in both Maths and Reading and on white males' Maths scores. They find no effect of limited English

(LE) students on the academic performance of both white and black native female students. Supporting these findings, Diette & Uwaifo Oyelere (2017) also report no evidence of negative peer effects of LE students on females and white students. However, they note that a 10% increase in the share of LE students is associated with a decline in reading scores of 0.013 standard deviations for males.

While the results indicate some negative effects of LE students, they also find that, after accounting for other factors, an increase in the share of Latin American (LA) students does not lead to negative peer effects on the achievement of native students. Their results, therefore, suggest that it is the language ability of peers that negatively affects natives in the classroom not racial or ethnic difference.

In contrast, Morales (2022) finds a positive impact of non-native students on the outcomes of native students. She demonstrates that moving from the 10th to the 90th percentile in cumulative exposure to non-native peers leads to increases in Math and Reading scores of 2.8% and 1.7% of a standard deviation, respectively.

The estimated non-linear impacts are also varied. Diette & Uwaifo Oyelere (2017) find that the negative effects of LE students on males are concentrated in the top third of prior achievement. In contrast, Morales (2022) identifies significant negative spillovers in English Language Arts (ELA) scores among low-achieving students, while high-achieving students experience positive spillovers. Additionally, Hanushek et al. (2009) show that higher-achieving Black pupils are more sensitive to the racial composition of their school, with the effects increasing monotonically across the initial achievement distribution and being significant at the five percent level only for the top two quartiles.

Studies using instrumental variables

Instrumental variables correct for endogeneity if they are correlated with changes in the composition of peer groups but uncorrelated with individual achievement. Fewer studies have used this methodology.

d'Este & Einiö (2021) examines the effect of Asian peers by leveraging a notable increase in birth rates within the Asian population, which is driven by the cultural belief that children born in the Chinese Year of the Dragon are thought to be luckier and more intelligent compared to those born in other zodiac years.[4] Their method takes advantage of the variation in the intensity of this fertility surge across different areas, based on the historical concentration of the Chinese population in various neighborhoods of New York City.

They present positive and significant first-stage coefficients to demonstrate that the interaction between the cohort dummy (equal to 1 if the individual was born in the Year of the Dragon) and the 1990 Chinese population share in a school's neighborhood, used as an instrument for the proportion of Asian students, meets the validity requirements.

d'Este & Einiö (2021) estimate that a 10-percentage-point increase in the share of Asian students reduces non-Asian pupils' Maths and English Language scores by 0.14 and 0.16 standard deviations. Contrary to Hanushek et al. (2009) findings, they find that the negative impact is mainly focused on Hispanic pupils.

d'Este & Einiö (2021) finds that the negative effects are most pronounced among students with lower Maths abilities. A 10-percentage-point increase in the share of Asian students results in about a 5.1 percentage point rise in the number of non-Asian students failing to meet Maths standards. For English Language, the decline in test scores is more widespread, affecting all ability levels: top-performing students experience a decrease, while there is a

[4] The Year of the Dragon lasts from February 5, 2000, to January 23, 2001.

smaller gain among those in the second-highest tier, indicating a shift of some students to lower performance levels. However, there is also a reduction in the number of students at the lowest performance level, suggesting that some lower-achieving students may benefit from a higher proportion of Asian peers in English Language classes.

Studies such as Geay et al. (2013) and Jensen & Rasmussen (2011) use instruments to analyse the impact of non-native peers.

Geay et al. (2013) exploit the increased demand for Catholic schools following the EU's expansion to include more Eastern European countries. They use an instrumental variable approach, leveraging the interaction between school type and the time trend after EU enlargement to instrument for the percentage of white non-native speakers in a given year group.[5] Jensen & Rasmussen (2011) use the concentration of immigrants in a broader geographical region as an instrument for the local concentration of immigrants. They argue that this instrument is valid because geographic mobility is constrained by factors such as employment and family ties, and the broader regional immigrant concentration does not directly impact child outcomes. They both show strong and positive first-stage coefficients, which convinces us about the strength of the instruments.

They both find negative OLS estimates, though, their IV estimates differ. Geay et al. (2013) estimate small and insignificant coefficients. Though they are insignificant they find a positive effect on Maths and a negative effect on English. This is logical because the differences between native and non-native speakers are generally smaller in mathematics compared to reading and writing. Therefore, if any positive spillover effects were to occur, they would be most likely observed in mathematics.

Jensen & Rasmussen (2011) IV estimates indicate that a 10 percentage-point increase in

[5] In May 2004, ten Central and Eastern European countries became members of the European Union. Initially, only the UK, Ireland, and Sweden permitted unrestricted movement for nationals from these new member states. The enlargement significantly increased the demand for Catholic schools, largely due to the high number of Polish families who practice this faith.

immigrant concentration reduces native pupils' Maths and Reading scores by 10.5 and 3.1 points respectively, though the findings are only significant for Maths scores.

Instrumental variables estimates generate local average treatment effects. As Jensen & Rasmussen (2011) use data on secondary school pupils from Denmark and Geay et al. (2013) uses data for primary school pupils in England, this could explain the difference in findings.

Studies using difference-in-differences

A small area of the literature, including Tumen (2021) and Contreras & Gallardo (2022), use difference-in-differences analysis by exploiting mass inflows of specific non-native groups.

Tumen (2021) analyses the effects of the influx of Syrian refugees into Turkey. Before January 2012, there were no Syrian refugees in the country, but their numbers have grown steadily since then.[6] Contreras & Gallardo (2022) explores the impact of both Venezuelan migrants (who speak the native language) and non-native-speaking migrants (primarily Haitians) on native students in Chile.

Tumen (2021) tackles the issue of refugees' endogenous location choices by employing an instrumental variable difference-in-differences (IV-diff-in-diff) approach.[7] He uses the variation in the refugee-to-population ratio across regions and over time in Turkey. To construct the instrumental variable, he weights the distances between Syrian source governorates and Turkish destination provinces based on the proportion of Syria's population living in each governorate prior to the conflict.

Contreras & Gallardo (2022) categorise a pupil as treated if they had no classmates at the grade level from Venezuela, Haiti, or other non-Spanish-speaking countries in 2016,

[6]Official statistics indicate that by the end of 2020, approximately 3.6 million Syrian refugees were registered in Turkey.

[7]An IV-diff-in-diff is a difference-in-differences method with continuous treatment, where the treatment variable is instrumented to address potential endogeneity issues.

but had at least one such classmate in 2018. A student is considered a control if they had no classmates from these countries at their grade level in both years. They interact the treatment variable with a dummy variable that is equal to 1 in the year following the migrant influx (2018) and 0 in 2016. This approach enables them to include individual fixed effects while ensuring that the treatment effect remains in the regression model.

Both studies justify the common trend assumption, suggesting that, without the inflow of specific non-native groups, the educational outcomes of the treatment and control groups would have evolved similarly over time.

Tumen (2021) shows that the Maths, Science, and Reading scores of Turkish 15-year-olds improved after the influx of Syrian refugees. The findings reveal that a one percentage point rise in the refugee-to-population ratio significantly boosts Maths, Science, and Reading scores by 0.048, 0.051, and 0.058 standard deviations, respectively. The impact is primarily concentrated in the lower half of the test score distribution, with similar effects observed for both males and females.

On the other hand, Contreras & Gallardo (2022) produce estimates to suggest that an increase in exposure to immigrants, decreases native test scores of 12-year-olds in both Maths and Reading by -0.046SD and -0.051SD respectively. The effects are larger for males and considerably greater when examining the migration of non-Spanish speakers compared with Spanish speaking (0.073SD compared with -0.056SD for maths and -0.099SD compared with -0.067SD for reading).

Tumen (2021) attributes their positive results to the labor market mechanism, where the influx of migrants filling lower-tier positions boosts competition in the low-skill sector. This, in turn, creates extra motivation for native workers to enhance their educational achievements.

Contreras & Gallardo (2022) argue that their results may be due to non-native speakers requiring more of the teacher's time. Therefore, the teachers spend less time focusing on the specific needs of native pupils.

Their contradictory findings could also be due to using different samples. The impact on younger aged pupils seems to be smaller.

Despite there being a lot left to debate within this literature, there are some interesting findings. Firstly, much of the literature finds the impact is larger for males. Secondly, pupils' Maths scores seem to be less impacted by peer group composition than English scores. Finally, there is some evidence to suggest that the negative impact of peer group composition could be explained by language difficulties.

4.2.6 Aspirations

As discussed above, the research on educational peer effects has focused almost exclusively on educational attainment. However, individuals' social networks might play a role in pupil aspirations, effort, perceptions about schooling returns, dropout decisions, and further education choices.

Whilst fewer studies have focused on these outcomes, five recent papers that have successfully analysed the impact of peers on educational outcomes beyond attainment are De Giorgi et al. (2009), Mora & Oreopoulos (2011), Mendolia et al. (2018), Dickerson et al. (2018) and Gagete-Miranda (2022).

Much of the research has explored how peers affect individuals' decisions regarding post-compulsory education. De Giorgi et al. (2009) investigated the influence of peers on college major selection, while Mendolia et al. (2018) examined how high school peers impact the likelihood of attending university and the chances of enrolling in a highly regarded university in England. Dickerson et al. (2018) focused on the choice of post-compulsory route taken, whilst Mora & Oreopoulos (2011) and Gagete-Miranda (2022), both analysed the impact on school dropout decisions.

Peers of Peers

The most common methodology used in this literature is instrumental variables. De Giorgi et al. (2009), Mendolia et al. (2018), Dickerson et al. (2018), and Gagete-Miranda (2022) all use peers-of-peers methodology.

Peer-of-peers methodology is argued to correct for endogeneity and reflection as each individual's peers are instrumented with their peers' peers , who have no direct interaction with the individual.

De Giorgi et al. (2009) use Italian data on a business school in one university. At the beginning of university, all students take the same course and attend lectures in randomly assigned classes for each course. De Giorgi et al. (2009) define peers as those students who attend at least four classes together. They then define peers of peers as the students who are in at least four classes of their peers.

They use admission test scores, high school final grades, and preferences for economics of the peers of peers as instruments for peers' college major choices. Results from first-stage regressions show that each instrument is significant at a 1% significance level.

Their results indicate that students are more likely to choose a major when many of their peers opt for the same field. While the OLS estimates are consistently insignificant, the IV estimates are significantly larger, suggesting that each additional peer selecting economics increases the probability of choosing economics as a major by approximately 7.4 percentage points.

They also explore the long-term effects of peer influence. If interactions with peers primarily involve the exchange of valuable information, then following peers' choices should result in better job placements. On the contrary, if students are influenced by their peers merely for social reasons, this could lead to poorer job matches in the labour market. By examining the proportion of peers who chose the same major as the student and comparing this with the student's relative ability, they find that peer-influenced students tend to have lower final grades and earn about 13% less than those driven by their own academic abilities. This suggests that choosing a major based on peer influence has consequences beyond just academic performance.

Mendolia et al. (2018) and Dickerson et al. (2018) improve on De Giorgi et al. (2009) approach by excluding each individual's own peers from the peers-of-peers' calculation. They both use the characteristics of the primary school peers of their secondary school peers, as an instrumental variable. Mendolia et al. (2018) and Dickerson et al. (2018) use the average ability, measured as KS3 test scores, whilst Dickerson et al. (2018) also use socio-economic background and aspirations to examine the impact on an individual's plan to stay in post-compulsory education, as well as their intentions to follow an academic rather than vocational pathway.

They both present positive and significant first-stage coefficients as well as large F stats, satisfying the validity of the instruments.

Mendolia et al. (2018) find that their OLS estimates show a positive correlation between the average quality of peers and an individual's test scores at ages 16 and 18. However, the IV estimates are notably higher, especially for performance at age 16, where a one standard deviation increase in the average KS3 score of peers is associated with about a 13% higher probability of taking A-levels. Despite this, the effect does not extend to university attendance. They find no significant impact of peer ability on the chances of enrolling in university or gaining admission to a prestigious higher education institution.

The findings are reinforced by Dickerson et al. (2018), who demonstrate that at age 16, there is evidence of positive peer effects, even after using instrumental variables. However, these effects are modest in size and are only statistically significant for boys.

Dickerson et al. (2018) further contribute to the literature by showing that, given an individual's intention to stay in post-compulsory education, peers appear to exert a stronger influence on their academic ambitions. Specifically, a 10 percentage point increase in the share of peers pursuing an academic route is associated with a 5 to 6 percentage point increase in the individual's likelihood of having similar academic aspirations.

Friends rather than peers

The peers-of-peers methodology assumes that every individual within a pupil's reference group is equally connected and exerts the same level of influence. However, in reality, individuals in social networks are connected in unique ways, and homophily has influences in the formation of friendships. There is a small number of studies that use data on nominated friends to determine the peer effect.

Mora & Oreopoulos (2011) were the first to adopt this methodology. By examining the impact of self-reported friends that are not reciprocated, they can control for the reflection problem. They discover that the influence of non-reciprocated peers' intentions to drop out is minimal and statistically insignificant. However, their analysis is based on the conditional independence assumption, which suggests that, given a set of observed variables, friendship groups are formed randomly. As previously discussed, this assumption is unlikely to hold, meaning their estimates are probably biased.

Gagete-Miranda (2022) builds on this approach by using data on the four closest friends of pupils in Brazil to examine the impact of peers on an individual's aspirations.

They model friendship formation based on pupils' similarities in pre-determined characteristics and their exogenous chances of interacting due to random assignment to classes. They find that a pupil is more likely to befriend another pupil if they both share their gender, race, and first-name initial. The identification strategy leverages the model's predicted relationships by using the characteristics of predicted friends-of-friends as instrumental variables for friends' aspirations, specifically defined as the probability of completing secondary education. They show strong first-stage coefficients with large F statistics.

The results show evidence of large, positive, and significant peer effects on aspirations. An extra friend finishing high school increases an individual's likelihood of finishing school by 6.62 percentage points. They show that the impact is more concentrated among low-income backgrounds, as they possibly rely more heavily on support from peers.

They also show that friends' college aspirations have a crucial impact on students' own college aspirations. An extra friend aspiring to a college degree increases a pupil's likelihood of also aspiring to it from 3.8 percentage point (5.6%) to 15.3 percentage point (22.5%), depending on the number of nominated friends.

4.2.7 Critical reflection

Despite methodological improvements, weakness still exist.

Validity of instrument

The robustness of observational studies such as Jensen & Rasmussen (2011), Geay et al. (2013), De Giorgi et al. (2009), Mendolia et al. (2018), Dickerson et al. (2018), d'Este & Einiö (2021), and Gagete-Miranda (2022) hinges on the strength of their instrumental variables. However, the effectiveness of these instruments can be uncertain, suggesting that endogeneity issues may still result in biased estimates.

Bramoullé et al. (2009) state that identifying peer effects through the use of peers-of-peers as instrumental variables is only feasible when distinct separations exist within peer networks. De Giorgi et al. (2010) rely on the assumption that students can only be considered peers if they attend classes together. The underlying assumption is that non-academic interactions, such as shared friendships and social activities outside of college, do not influence the choice of a college major. This assumption is debatable as studies, for example, Muñoz-Bullón et al. (2017) show that participation in organised sports is linked to improved academic performance among university students, though the mechanism for this effect is not clear.

De Giorgi et al. (2010) do try to control for this limitation by constructing placebo peer groups by assigning students to hypothetical classes randomly. They find no indication of significant social interactions with the magnitude of the point estimates being small and close to zero.

As mentioned above, Mendolia et al. (2018) and Dickerson et al. (2018) do improve on De Giorgi et al. (2010) approach by using the primary school peers of a pupil's secondary school peers, excluding one's own earlier peers to create a clear divide in network groups. Mendolia et al. (2018) and Dickerson et al. (2018) do find smaller estimates than De Giorgi et al. (2010).

A further concern regarding the peers of peers instrument is that pupils might not follow the behaviour of the average person in their class. Pupils may be more likely impacted by a "shining light" or "bad apple". The presence of significant outliers means that using the average peer behaviour as an instrument might not accurately reflect the typical peer effect, leading to biased estimates.

Gagete-Miranda (2022), who argues that it is close friends that impact performance, tries to control for the issues highlighted above by including classroom-fixed effects in addition to school-fixed effects. With students' rankings and competitive dynamics within the classroom remaining constant, the exclusion restriction should be valid. They find no change in their results.

Despite the possibility that pupils might be impacted by a small circle of peers rather than their full network, there is potential for measurement error when using friends rather than peers. Pupils, depending on their age, tend to change friendship groups, and therefore measuring their connections is a challenge. Gagete-Miranda (2022) encounters a ceiling effect because pupils were limited to nominating only four friends. They indicate that approximately 20% of pupils may have been affected by this ceiling effect, as they nominated the maximum four friends, making it unclear if there were additional friends they wished to include.

Gagete-Miranda (2022) also provides an analysis based on a subset of students who identified three or fewer friends. With this limited sample, it is easier to map all student relationships precisely, minimising the risk of overlooking connections. The findings from this sample are nearly identical to the original results.

Omitted variables

The main threat to identification within this literature is unobserved non-random selection into schools. The most common methodology used in the literature is high dimensional fixed effect estimation which controls for time-invariant selection.

Hanushek et al. (2009), Diette & Uwaifo Oyelere (2017), Maestri (2017), and Morales (2022) all use between cohort variation to estimate peer effects. However, this approach faces potential challenges due to variations between adjacent cohorts, such as differences in mobility, changes in teacher or school attributes, and disruptions within the school or neighborhood. These factors could be systematically connected to both racial composition and academic performance, potentially resulting in biased estimates.

Mobility-induced changes in racial composition pose significant challenges. Hanushek et al. (2004) shows that Black pupils are more likely to switch schools compared to white pupils, which can lead to disproportionate impacts on year-to-year variations in school racial composition. Additionally, evidence indicates that students who move tend to have lower prior achievement. As a result, increased Black enrollment in schools may be associated with lower average achievement for Black students, potentially biasing estimates. Furthermore, mobility can affect class sizes, creating another potential source of spurious correlation between achievement and the proportion of ethnic minorities.

Hanushek et al. (2009) tackle these concerns by examining the interaction between mobility and the proportion of Black students to understand how movers—who are more likely to face disruptions over time—might impact the estimates. They divide students into three categories: (1) those who remain at the same school, (2) those transitioning from elementary to middle school within the same district (a structural move), and (3) those relocating to a new district (a family move). Their hypothesis suggests that students who stay at the same school are less likely to experience changes in their social networks, benefiting from the stability of their friendships.

The findings show that mobility does not significantly influence the results, as the coefficients for the interaction between the proportion of Black pupils and school movers are much smaller and less significant compared to those for interactions with non-movers.

Incorporating teacher and school attributes—such as class size, the percentage of first-year teachers, and school mobility rates—does not change the estimated coefficients. This is notable given that both class size and the ratio of inexperienced teachers are important factors influencing academic performance.

Another omitted factor that may vary over time is family related such as parents' jobs, income, and attitudes. It can be argued that these factors impact both the ethnic composition of a pupil's school and their academic achievement, therefore leading to biased estimates.

A school might experience a trend in ethnic diversity that correlates with changes in other factors (such as new waves of immigrants), which could impact school achievement. Additionally, parents may be aware of the increasing ethnic diversity in the school and might adjust their choice of school for their children accordingly.

Figlio et al. (2024) and Green & Iversen (2022) uses variation in sibling exposure to non-natives and finds slightly larger estimates than studies without family-fixed effects.

However, this approach has its challenges. Their method is based on the assumption that families choose schools independently of their children's specific characteristics. If parents choose to enroll their highest-performing child in a school with a lower immigrant population, the estimated coefficient could be underestimated. Alternatively, if parents believe that being around immigrants adversely affects their children's performance, they might opt to send their lower-performing child to a school with fewer immigrants, potentially resulting in an inflated coefficient. Figlio et al. (2024) argues that there is minimal variation within families.

Another area not controlled for is segregation within schools. When a large number of students with limited English proficiency attend the same school, they might be placed in specialised classes for extra support. This could lead the school to offer more targeted resources to all students. If this is the case, the results might just indicate that schools with fewer immigrants have fewer resources available.

Figlio et al. (2024) attempts to explore this issue by analysing aggregate school-level data on classroom distribution. They present evidence that counters the earlier hypothesis by demonstrating that the estimated coefficient is consistently higher in schools with lower segregation. However, their analysis is constrained by the use of only aggregate-level data, which prevents them from examining more detailed individual-level classroom effects.

Generalisability of findings

There is limited consensus within this literature. Depending on the country, time frame, and methodology used, the estimated impact can vary considerably.

If the OLS estimates reflect a downward bias due to high-ability individuals choosing not to live in areas with many immigrants and ethnic minorities due to hostile attitudes, then correcting for this bias should increase the estimated impact (Dustmann & Preston 2001).

De Giorgi et al. (2009) and Mendolia et al. (2018) both find IV estimates larger than OLS. However, Dickerson et al. (2018) using the same estimation strategy find IV estimates that are sometimes bigger and sometimes smaller than the LPM . Maestri (2017) and Figlio et al. (2024) using fixed effects find estimates larger than OLS whilst Chareyron et al. (2021) find estimates smaller than LPM. These differing findings are not explained.

In general, the literature presents larger estimates in studies using instrumental variables. This could be due to the estimation strategy producing local average treatment effects (LATE). Gagete-Miranda (2022) shows that pupils from low-income backgrounds are more impacted by peers. If the random shock to peer group composition is more likely to change peer groups for pupils from low-income backgrounds, then this could explain the larger estimates.

The results also differ based on the country being examined. Schnepf (2007) explores differences in educational performance between immigrants and native students across ten countries. In Australia and Canada, a higher concentration of immigrants appears to improve students' educational results. In contrast, this effect is not significant in countries like the Netherlands, Sweden, the UK, and the US. Additionally, in Switzerland, Germany, New Zealand, and France, a high concentration of immigrants seems to have a negative impact on achievement. However, it is crucial to recognise that the analysis is based solely on simple

OLS, which may lead to bias.

Another factor that limits the generalisation of findings, is that non-natives and ethnic minorities are a heterogeneous group. Studies that analyse the impact of non-natives define their population of interest differently. Morales (2022) analyses the impact of self-reported refugees whilst Jensen & Rasmussen (2011) look at first- and second-generation immigrants. Other studies, for example, Tumen (2021) examine the impact of specific groups of immigrants such as Syrian refugees. Others like Geay et al. (2013) and Contreras & Gallardo (2022) focus on language spoken at home. These studies obtain varying findings from large positive impacts (Tumen 2021) to small insignificant estimates (Geay et al. 2013) and negative impacts (Contreras & Gallardo 2022). Much of the research on ethnic minorities analyses the impact of black and white pupils with limited research carried out on other ethnicities. Studies which focus on one group might be less useful from a policy perspective due to the diverse nature of many countries.

Also, most studies look at the impact of the minority group on the majority group without contemplating the impact of the majority group on the minority. This is also important from a policy perspective. However, the studies which do look at the impact from both sides do not draw similar conclusions. Jensen & Rasmussen (2011) suggest that immigrant concentration in a school is more important for natives than for immigrants whilst Szulkin & Jonsson (2007), find that it is more important for immigrants.

Focusing on ethnicity, Hanushek et al. (2009) estimates that a higher percentage of Black classmates is associated with lower achievement among Black students. In contrast, this effect is much smaller and generally insignificant for white and Hispanic students. The very limited evidence makes it difficult to provide a convincing argument.

4.2.8 This research's contribution

Despite improvements in econometric analysis and large representative datasets, researchers are divided on the impact of peer characteristics.

This research contributes to the literature by analysing the impact of ethnic diversity in a classroom on pupils' aspirations in England. To the best of our knowledge, there are no other studies that answer a similar question.

We have highlighted that there is a range of literature that looks at the impact of ethnic diversity on pupil achievement outcomes. Also, a separate strand of the literature looks at the impact of peers on an individual's aspiration. However, these two literature strands have not yet been brought together.

This research, therefore, aims to make three contributions. Firstly, we provide the first evidence on the impact of peers' ethnicity on individuals' perceived likelihood of attending university. Secondly, we focus on the impact of ethnic minorities. The literature has mainly examined the impact of refugees or immigrants with native language challenges. Finally, we add to the literature using instrumental variables by creating a unique instrument.

Analysing the relationship between peers' characteristics and individual outcomes is impacted by methodological challenges. Research has varied considerably in terms of methodologies, countries, and time spans, generating differing findings. In 16 studies over 10 years, results have ranged from finding positive effects to statistically insignificant relationships to large negative relationships. Therefore, it is important to continue adding to this evidence base.

4.3 Data

The data used in this research combines the Millennium Cohort Study (MCS) with the National Pupil Database (NPD).[8] For full information on these datasets see Section 2.3 and 3.3.

4.3.1 Sample

In this research we use data from wave 6 of the MCS, which took place when the cohort members were age 14.[9] We restrict our sample to children living in England, as the NPD is only available for children who attend school in England. For our baseline model we also restrict the sample to white pupils as the research question specifically looks at the impact of ethnic minority peers on white pupils' perceived likelihood of attending university. Due to the limited sample size, we are unable to examine the impact of white pupils on ethnic minorities.

The sample used in this study covers 3,607 white pupils who report their likelihood of attending university at age 14 as well as having other information including parental education and income.[10] We also require that their survey data has been successfully matched to school level information on peers' ethnicity.

[8]Data reference: University College London, UCL Institute of Education, Centre for Longitudinal Studies, Department for Education, 2023, Millennium Cohort Study: Linked Education Administrative Datasets (National Pupil Database), England: Secure Access, [data collection], UK Data Service, 2 nd Edition, SN: 8481, DOI: http://doi.org/10.5255/UKDA-SN-8481-2

[9]Wave 6 was carried out between January 2015 and April 2016 when the cohort members were on average aged 14.

[10]The Millennium Cohort Study is nationally representative. Whilst 11,859 individuals were surveyed at age 14, only 4,954 English white pupils reported their university likelihood. The further reduction in the sample is due to missing data in the control variables. Despite being a relatively small sample, it is representative. We discuss the implications of attrition in the MCS in Chapter 2 Section 2.5.5.

4.3.2 Measure of university likelihood

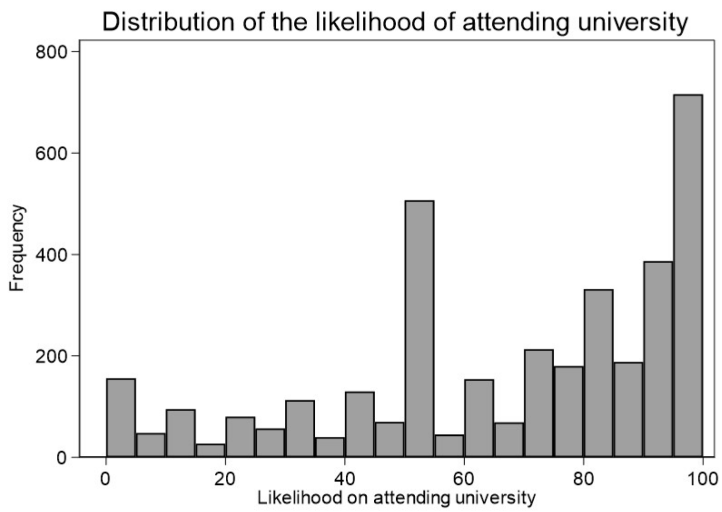
The outcome of interest in this research is the perceived likelihood of attending university. We focus on the reported likelihood of attending university as the literature to date discussed in Section 4.2, has mainly focused on the impact of ethnic minorities on white pupils' academic achievement. We believe it is important to assess the impact on further educational goals as they provide pupils with a sense of direction and purpose.

As our measure of the perceived likelihood of attending university, we use a question asked to cohort members in wave 6. The cohort members were asked how likely it was, on a scale of 0-100%, that they would go to university.

Whilst the question on university likelihood is also asked in wave 7 of the MCS when the members are 17, we do not examine the impact of peers' ethnicity on this measure. Pupils in England decide whether to take an academic or vocational route at age 16, this suggests that by age 17, the desire to attend university has already been determined.

Figure 4.1 illustrates the distribution of university likelihood at age 14. Similar to many subjective probability questions, De Bruin et al. (2000) found that responses tend to cluster at multiples of 5 and 10, with a notably high percentage of answers given as '50%.' This indicates that the question is affected by different forms of reporting behaviour, such as rounding and focal responses. Kleinjans & Soest (2014) who investigate how individuals respond to survey questions about their personal beliefs or expectations, particularly when asked to assign probabilities to uncertain events, find that these behaviors can lead to biased data and missing responses, posing challenges for accurately interpreting results. They discuss the use of Tobit models as one of the potential methods to handle issues with subjective probability data. In their analysis, they indicate that using Tobit models allows for the estimation of the underlying distribution of subjective probabilities while accounting for the censoring at the bounds. This can help in deriving more accurate measures of individuals' true subjective probabilities but it might not fully address the rounding behavior. They present evidence demonstrating that estimating a tobit model yields signs and significance levels closely resembling those of models that fully account for reporting behaviour. They conclude that overlooking reporting behaviour results in only modest biases in the estimated means and standard deviations of the actual probabilities. Based on these finding we use as OLS estimation. We also estimate a Tobit model to test the sensitivity of our findings.

Figure 4.1: Distribution of university likelihood reported at age 14



4.3.3 Measure of peers' ethnicity

The NPD provides school level information on the number of pupils from different ethnic groups. We use this information to calculate the percentage of pupils within a school who have a non-white ethnicity. We define a pupil as having a non-white ethnicity if they identify as Asian, Black, mixed raced or other.[11]

In English secondary schools, pupils are placed with different peers for each subject, allowing

[11]Asian pupils include Pakistani, Indian, Bangladeshi, other Asian and Chinese. Black includes Black African and Black Caribbean. Mixed raced include White and Black African and White and Asian. We remove mixed raced from our definition of non-white as a robustness check, see Section 4.5.8

them to interact with a much larger group of classmates compared to primary school. Despite this, we recognise that peer effects may be larger from interactions within smaller groups, such as close friends, rather than from the wider school environment. Unfortunately the available data does not provide information on peer ethnicity at the year group level or information on friendship groups.

To address this concern, we make two points. Firstly, the NPD reports the ethnic makeup of a school every year. The variation of non-white pupils within a school across years is small, indicating that the ethnic make-up of each year group within a school is similar. Secondly, as a robustness check, we make use of a question in the MCS asked to the cohort members about their friends. In wave 6, cohort members are asked "How many of your friends are from the same ethnic group as you?" This is discussed further in Section 4.5.8.

4.3.4 Conditioning variables

The MCS linked with the NPD provides us with rich background information on the cohort member, household and school attended. We include these in the model to control for factors that may impact upon an individual's reported likelihood of attending university and ethnicity of peers in order to identify the impact of peers' ethnicity on the individual's perceived likelihood of attending university. See Table 4.2 in Section 4.5.1 for the full summary statistics.

We consider the individual characteristics (gender, month of birth, KS2 score[12]); household characteristics (number of siblings, real weekly equivalent income[13] , and local deprivation index); main parent's characteristics (age and level of education and economic activity sta-

[12]The KS2 score is used as a proxy for ability.

[13]The equivalent income is the income of the household taking into account the number of people in the family and assigning weights. The one provided in the MCS follows the OECD equivalence scale, which assigns a value of 1 to the first household member, of 0.7 to each additional adult, and of 0.5 to each child.

tus)[14] ; and school characteristics (number of students enrolled in the school, average KS2 performance of the current KS4 pupils[15] , number of pupils on free school meals).

4.4 Method

4.4.1 OLS estimation

The data allows this research to observe university likelihood at age 14 as well as the ethnic makeup of the school attended. This analysis aims to identify a causal effect of non-white peers on white pupils' university likelihood. As previously discussed in Section 4.2, identifying causality between peers' ethnicity and pupils' university likelihood faces methodological challenges. There are two main threats to causality: (i) omitted variable bias, and (ii) self-selection in to peer groups. Section 4.3 discussed the controls we include to minimise omitted variable bias. This section will outline how this research will try to correct for self-selection.

The analysis starts by estimating an ordinary least squares regression (OLS) in which we regress university likelihood of white pupils on the proportion of peers who have a non-white ethnicity and the covariates.[16] The model takes the following form:

$$UniversityLikelihood_i = \beta_0 + \beta_1 Non-WhitePeers_i + \beta_2 Child_i + \beta_3 Family_i + \beta_4 School_i + \epsilon_i \tag{4.1}$$

Where $UniversityLikelihood_i$ is the self-reported likelihood of attending university for indi-

[14]The main parent is identified as the parent who responds to the survey about the cohort member. We also run a model where we control for both main and second parent characteristics and find no difference in the estimated coefficients.

[15]We use this as a proxy for school quality.

[16]We assume that university likelihood depends linearly on peer ethnicity. This assumption is tested in Section 4.5.3.

vidual i. $Non-WhitePeers_i$ is the proportion of pupils within the individual's school who have a non-white ethnicity. $Child$ is a vector of individual characteristics including gender, month of birth and KS2 achievement. $Family$ is a vector of household and parental characteristics including the local deprivation index, number of siblings, real weekly equivalent income, main parent's highest education level, labour market status and age at birth of child. $School$ is a vector of school level characteristics including number of pupils enrolled in the school, average KS2 performance of the current KS4 pupils and the number of pupils on free school meals.

The main interest of this research is the estimation of β_1, which is the effect of non-white peers on university likelihood. To interpret β_1 as the causal effect, we require independent variation in the proportion of non-white peers, meaning the zero conditional mean assumption must hold, $E(\epsilon_i|Child_i, Family_i, School_i) = 0$. Due to the endogeneity of the ethnic makeup of a school, it can be argued that this assumption may not hold.

The school the pupil attends is largely driven by parental location choice. Ethnic minorities are not spread equally across the country and tend to be very geographically concentrated. Individuals who attend a school with a high percentage of ethnic minorities may differ in both observable and unobservable ways to those who attend schools with low levels of ethnic minorities. These differences could be driven by child and parental factors. Parents with similar characteristics tend to select into the same areas. As discussed in Section 4.2, Dustmann & Preston (2001) found, using data from England, that individuals who hold negative views towards ethnic minorities tend to avoid living in areas with a high concentration of those groups. Betts & Fairlie (2001) and more recently, Cascio & Lewis (2012) provide evidence in support of 'white flight', where white pupils move to private schools as a response to increasing levels of ethnic minorities. Ethnic minorities, specifically immigrants, are more likely to settle in areas with a lower-than-average level of education (Dustmann & Preston 2001). However, ethnic minority parents tend to have high expectations for their children and therefore may select higher-quality schools (d'Este & Einiö 2021). Many of these factors may not be observable and may influence the child's reported likelihood of attending university, biasing the estimates. The direction of the bias is not straightforward. Take school quality for example. Whilst this is likely to increase the reported likelihood of attending university it is not clear whether this would increase or decrease the number of ethnic minority peers. As discussed in Section 4.2, much of the literature to date has suggested a downward bias mainly driven by high-ability individuals choosing not to live in areas with many ethnic minorities (Dustmann & Preston 2001). In this case we can interpret the OLS results as a lower bound of the association between ethnic minority peers and the perceived likelihood of attending university.

4.4.2 Instrumental variable

To check whether the estimated effect is subject to bias, we exploit exogenous variation in peers' ethnicity by using an instrumental variable. We argue that the current geographical distribution of ethnic minorities in England is driven by labour shortages post World War 2 and government backed requirements specifically in the health care industry. Appendix Figure C.1 provides a map of the geographical distribution of ethnic minorities in England.

The 1948 British Nationality Act allowed all Commonwealth citizens to obtain British passports and work in the UK.[17] Britain encouraged mass immigration from the Commonwealth countries after the Second World War due to severe labour shortages, especially in the newly created National Health Service (NHS)[18].

Hospitals in Britain were dealing with labour shortages well before the establishment of the NHS and nurse shortages had been discussed in several government inquiries.[19] The national

[17]See Home Office (2020) for an overview of the history of nationality law.

[18]The NHS was established in 1948.

[19]See Parliament. House of Commons (1942) where Mr. Ernest Brown the Minster of Health discusses

post-war labour shortage had only made the problem worse. By 1948, there were 54,000 nursing vacancies (Snow & Jones 2011). In 1949, the Ministries of Health and Labour, along with various healthcare representatives such as the Colonial Office, the General Nursing Council, and the Royal College of Nursing, initiated campaigns to recruit hospital staff directly from the West Indies (NHS 2023). Senior NHS personnel from Britain traveled to the West Indies for recruitment, and job vacancies were advertised in local newspapers. By 1955, official nursing recruitment programs had been established in 16 British colonies and former colonies (Snow & Jones 2011). The NHS, which became Britain's biggest employer in 1961, recruited thousands of workers from the Commonwealth. By 1968, Commonwealth migrant filled approximately a third of student nurse and midwife roles (Babikian 2021).

It was not just nurses that were required from overseas, the labour shortages also drove the first mass wave of junior doctor recruitment from India, Pakistan, Bangladesh and Sri Lanka. The government was also involved in the recruitment of transport workers.

Additionally, labour shortages were not the only reason many individuals came to Britain. The Caribbean islands economy, underdeveloped by Britain, faced high levels of unemployment. Additionally, the partition of India and Pakistan and the civil war in Cyprus led many people to flee and seek a better opportunities in the UK. These individuals arrived in the country and undertook employment as carpenters, typists, tailors, machinists, domestic servants, etc.[20]

We choose the percentage of individuals employed as nurses in the Local Government District (LGD) as our instrument for two main reasons. Firstly, it was a government backed recruitment drive where immigrants were placed in the areas with the shortages. Secondly, nurses were needed across the country whereas some of the other professions were geographically concentrated. Once individuals arrived in Britain, they were dispersed to their appointed

the nurse shortages.

[20]We use these additional occupations as robustness checks. See Section 4.5.8.

hospitals all over the United Kingdom where they lived in the Nurses' Homes attached to the hospitals.

Until 1986, there were two nurse training programmes: State Registered Nurse (SRN) qualification and State Enrolled Nurse (SEN) qualification.[21] Most individuals arriving from the Commonwealth, were placed on the SEN course. It has been suggested that few were accepted on the SRN course despite possessing the required qualifications, due to racial discrimination.

In 1951, England was still faced with large labour shortages. We therefore assume that areas that had low levels of nurses in 1951 were the areas with the highest shortages. In the following years, these shortages were filled by individuals from the commonwealth. We argue that it is these areas that have higher numbers of ethnic minorities today. This is due to that fact that in the following years, there was growing public and political unease regarding the impact of migration, leading to numerous changes to the 1948 British Nationality Act. These changes focused on the entry of dependents and family members of those already in the United Kingdom meaning that individuals entering the country in later years normally located in areas where their family members already were. Additionally, Britton et al. (2021) provides evidence to show that ethnic minorities are less likely to move areas, and the effect of higher education on mobility is much weaker. After their nurse training, many individuals stayed where they had been placed. Most of them could not get onto the SRN course, and therefore could not get promoted. Many felt unable to return to their home country as the SEN qualification was not recognised in the Caribbean. In addition, it was very difficult for them to move to other places in the United Kingdom due to difficulty finding accommodation. The infamous 'No Irish, no Blacks and no dogs' signs have become symbolic of the wave of xenophobic sentiment that arose in response to the influx of people

[21]The State Registered Nurse (SRN) and State Enrolled Nurse (SEN) qualifications represented different levels of nursing training and responsibilities. The SRN qualification required three years of training, focused on medical and surgical nursing skills as well as taking on leadership roles. In contrast, the SEN qualification was a two year course covering the fundamentals of nursing care with limited clinical responsibilities.

who came to answer Britain's call for workers. The response also resulted in the 'colour bar', an informal discriminatory practice of the time whereby people of colour were denied jobs, housing and services or spaces, such as pubs, had segregated access.

Appendix Figure C.2 shows a map of the geographical distribution of nurses in 1951.

4.4.3 Data on nurses

The occupation data comes from the 1951 census which has been computerised by the Great Britain Historical GIS Project.

The 1951 census data is recorded at the local government district level (LGD). Since then, the local government structure in England has undergone significant changes meaning that LGDs are not directly comparable to current geographical areas. The MCS records the middle layer super output area (MSOA) the cohort member lives in at each wave. Using centroid mapping, which is where if the centre of the MSOA falls within the LGD it is assigned to that LGD, we match LGDs to MSOAs.

4.4.4 Two-stage least-squares (2SLS) regression approach

We use a two-stage least-squares (2SLS) regression approach to first estimate the proportion of non-white peers as a function of nurses in 1951, net of child, family, and school characteristics. The predicted proportion of non-white peers is then forwarded to a second-stage regression to estimate the unbiased LATE of non-white peers on university likelihood. The first stage equation takes the following form:

$$Non - WhitePeers_i = \beta_0 + \beta_1 Nurses_i + \beta_2 Child_i + \beta_2 Family_i + \beta_4 School_i + \gamma_i \quad (4.2)$$

Where $Nurses_i$ is the percentage of employed individuals in individual i's local government region who were nurses in 1951.

The second-stage equation takes the following form:

$$UniversityLikelihood_i = \beta_0 + \beta_1 Non - \hat{W}hitePeers_i + \beta_2 Child_i + \beta_3 Family_i + \beta_4 School_i + \epsilon_i \quad (4.3)$$

where $Non - \hat{W}hitePeers_i$ is the predicted proportion of peers who are non-white based on the first stage.

The IV strategy requires that three assumptions be met. First, the instrument must be relevant, meaning that the proportion of nurses in the local area in 1951 is highly predictive of the exposure to non-white peers. Second, it must be exogenous, meaning it is not correlated with the error term in the explanatory (second-stage) equation. Finally, the instrument must affect university likelihood only through its effect on the proportion of non-white individuals within a school and not through any other pathway. We discuss the second and third assumption together.

The first assumption is easily tested and as shown in the first stage results (Appendix Table C.5), holds true in our models. The second and third assumption is untestable meaning we cannot fully rule out this possibility, though we attempt to minimise it by adjusting for an extensive array of covariates.

The proportion of nurses in 1951 is a historical variable, relatively distant in time from 2014, making it less likely to have a direct effect on contemporary university choices.

We also argue that unobserved historical factors and endogenous policy responses are likely to have reduced impact over time and are therefore unlikely to have a persistent and direct impact on university choices in 2014. Additionally, we control for the index of multiple deprivation in 2014 to control for persistent socioeconomic factors.

4.5 Results

4.5.1 Descriptive Statistics

This research focuses on the impact of non-white peers on white pupils' self-reported likelihood of attending university. To examine the difference in reported likelihood of attending university across ethnic groups, we focus on an initial sample of 5,340 individuals who report university likelihood at age 14. 67.5% are white, 20.8% are Asian, 5.9% are Black, 4.8% are mixed raced and 1.0% classify themselves as another ethnicity.[22]

The motivation for this research, as set out in Section 4.1, is that ethnic minorities have higher university aspirations than white individuals. Table 4.1 presents the average reported likelihood at age 14 of attending university for each broad ethnic group.[23] Ethnic minority pupils are significantly more likely to report higher likelihoods of attending university than white pupils at age 14.

[22]This sample is representative of the UK population. See Office of National Statisitcs (2023B) for data on the population of England.

[23]See Table C.1 in the appendix for reported university likelihood across a wider range of ethnic groups. All groups report a higher likelihood of attending university than white pupils.

Table 4.1: Average reported likelihood of attending university at age 14.

Ethnicity	Probability of attending university (age 14)
White	65.75
Asian	80.49
Black	84.24
Mixed ethnicity	74.61
Other ethnic group	77.97

To examine the descriptive relationship between university likelihood and peers' ethnicity, we restrict the sample to white pupils only. The sample is now comprised of 3,607 white individuals who report university likelihood at age 14.

Descriptive statistics for this sample are shown in Table 4.2 [24][25] The average percentage of peers who have a non-white ethnicity ranges from 0 to 96.8%. The average is 8.86%.

[24]Standard deviation in parentheses.

[25]Parental education is measured in levels. See Table C.2 in appendix.

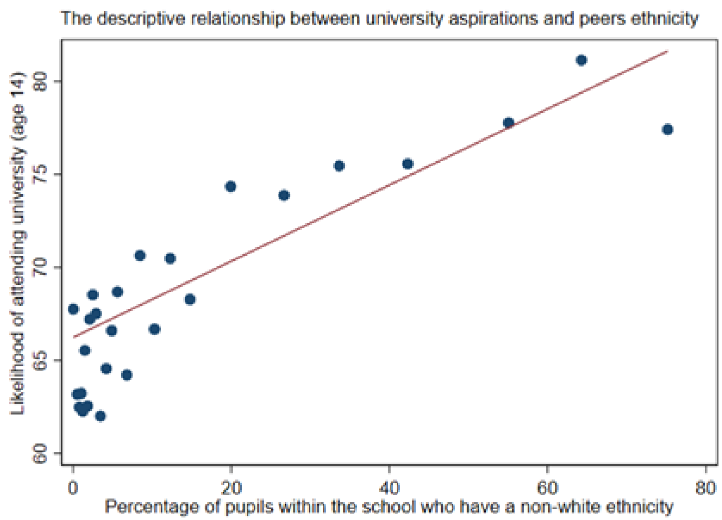
Table 4.2: Summary statistics for covariates

Variable	Mean
University likelihood age 14	65.75% (28.99)
Percentage of peers who have a non-white ethnicity at age 14	8.86% (14.84))
Female	0.51% (0.50)
KS2 score (Standardised) (Proxy for ability)	0 (1.00)
Main parent employed	0.69 (0.46)
Main parent self-employed	0.10 (0.29)
Main parent out of the labour market	0.15 (0.36)
Main parent unemployed	0.01 (0.11)
Main parent employment missing	0.05 (0.23)
Main parent education L1	0.13 (0.33)
Main parent education L2	0.39 (0.49)
Main parent education L3	0.10 (0.30)
Main parent education L4	0.26 (0.44)
Main parent education L5	0.03 (0.18)
Main parent other qualification	0.02 (0.12)
Main parent no qualification	0.07 (0.26)
Equivalised household income	458.31 (169.56)
Number of siblings	1.35 (0.98)
Main parent age at birth of cohort member	29.49 (5.58)

Most deprived decile	0.07 (0.26)
10-20%	0.08 (0.27)
20-30%	0.09 (0.29)
30-40%	0.08 (0.28)
40-50%	0.11 (0.31)
50-60%	0.10 (0.30)
60-70%	0.10 (0.30)
70-80%	0.11 (0.31)
80-90%	0.11 (0.32)
Least deprived decile	0.12 (0.33)
Total number of pupils enrolled full time	1,015 (39.62)
KS2 average point score of the cohort at the end of KS4 (Proxy for school quality)	28.68 (1.84)

To conclude the descriptive analysis, Figure 4.2 documents the association between the reported likelihood at age 14 of attending university and the percentage of peers who have a non-white ethnicity. The correlation is positive. As the percentage of non-white peers increases the average reported university likelihood for white pupils also increases. For white pupils who have less than 50% exposure to non-white peers, the average reported likelihood at age 14 of attending university is 65.5%. For white pupils who have over 50% exposure, the average reported likelihood of attending university is 71.9%.

Figure 4.2: Descriptive relationship between the percentage of peers who have a non-white ethnicity and perceived university likelihood for white pupils (age 14).



Notes: Correlation between the percentage of non-white pupils and university aspirations for white pupils at age 14. The percentage of non-white peers is split in to 36 equally sized groups. The average university aspirations is calculated for each of the categories. The fitted line is taken from a simple linear regression of university aspirations on the percentage of non-white peers.

4.5.2 Regression Results

Table 4.3 contains the OLS estimates for university likelihood from the model specified in equation (4.1). The OLS results indicate a positive and statistically significant association between peer ethnicity and perceived likelihood of attending university. The association is small in magnitude, with a 1 percentage point increase in the proportion of non-white peers being associated with a 0.08 percentage point increase in the reported likelihood at age 14 of attending university.

Table 4.3: Impact of ethnic minority peers on reported university likelihood (OLS estimation)

	(University likelihood)
Non-white	0.0812***
	(0.0301)
N	3607
R^2	0.12

Notes: Standard errors in parentheses. The dependent variable is the reported likelihood at age 14 of attending university (provided as a percentage). Non-white is the percentage of peers who have a non-white ethnicity. Controls include gender, a proxy for ability, month of birth, main parent's age at birth, number of siblings, main parent's education and labour market status, household equivalised income, the local deprivation index, the number of pupils enrolled in the school and a proxy for school quality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Turning to the IV results, the instrument performs well.[26] The first-stage F statistic (Montiel Olea and Pflueger F statistics) is 59.4.

The first stage estimates (see Appendix Table C.5) suggests that a 1 percentage point increase in the percentage of nurses in the local government district in 1951, reduces the proportion of non-white peers in 2014 by 1.67 percentage points.

As discussed in Section 4.4.2, in 1951 there was a shortage of nurses. Whilst government-backed recruitment of nurses from the commonwealth had begun in 1949 there were still significant shortages in 1951. As nursing is a profession required across the country, the 1951 census data shows where the shortages were. If the local government district had high numbers of nurses in 1951, they had less labour shortage and were therefore assigned few of the government requirements from overseas. Britton et al. (2021) provided evidence to show that ethnic minorities are less likely to move areas. We therefore expect to see fewer ethnic minorities in these areas today.

The second-stage IV results, shown in Table 4.4, are larger in magnitude than the OLS

[26]See the robustness checks for further discussion on the strength of the instrumental variable.

estimates.[27] This suggests a downwards bias in the OLS estimates. This is consistent with the literature, specifically with studies that use instrumental variables (De Giorgi et al. 2009, Mendolia et al. 2018). The IV results indicate that on average, increasing the percentage of non-white peers by 1 percentage point, increases white individuals' reported likelihood at age 14 of attending university by 0.46 percentage points. This result aligns with the work of Dickerson et al. (2018), who also analysed data from England and finds that a 10 percentage point rise in the proportion of an individual's peers aspiring to pursue an academic path is associated with a 5 percentage point increase in the individual's likelihood of having similar aspirations.[28]

Table 4.4: Impact of ethnic minority peers on reported university likelihood (Just-identified IV model)

	(University likelihood)
Non-white	0.4589***
	(0.1514)
First Stage F	59.40
N	3607

Notes: The dependent variable is the reported likelihood at age 14 of attending university (provided as a percentage). Non-white is the percentage of peers who have a non-white ethnicity. Montiel Olea and Pflueger F statistic are shown. Controls include gender, a proxy for ability, month of birth, main parent's age at birth, number of siblings, main parent's education and labour market status, household equivalised income, the local deprivation index, the number of pupils enrolled in the school, and a proxy for school quality. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5.3 Functional Form

The simplifying assumption made in the baseline model is that there are neither diminishing nor increasing benefits to the proportion of non-white peers.

[27] Full model second stage results are shown in Table C.6 in the appendix

[28] Table C.3 presents the IV estimates from a just identified model when controls for both parents are included. Table C.4 presents the IV estimates from a just identified tobit model. The estimated impact is consistent across the different models.

The next stage of the analysis explores the extent to which the proportion of non-white peers has a non-linear effect on university likelihood. As ethnic minorities are not equally spread across the country, some white pupils will be exposed to large numbers of peers from ethnic minority backgrounds whilst others will have very few non-white peers. To examine whether our estimates are driven by these differences we include a quadratic of the proportion of non-white peers in the OLS estimation.[29] Table 4.5 provides evidence to show that as the proportion of non-white peers increases, the impact on perceived likelihood of attending university decreases. The coefficient on the quadratic term is negative, though small and statistically insignificant. As we argue that the OLS estimates are biased downwards, we cannot rule out a non-linear relationship. To determine whether there is an optimal level of diversity within a school, further analysis is therefore required.

Table 4.5: Non-linear impact of non-white peers on university likelihood (OLS estimation)

	(University likelihood)
Non-white	0.2341***
	(0.0857)
Non-white^2	-0.0021
	(0.0013)
N	3607

Notes: Standard errors in parentheses. The dependent The dependent variable is the reported likelihood at age 14 of attending university (provided as a percentage). Non-white is the percentage of peers who have a non-white ethnicity. Controls include gender, a proxy for ability, month of birth, main parent's age at birth, number of siblings, main parent's education and labour market status, household equivalised income, the local deprivation index, the number of pupils enrolled in the school and a proxy for school quality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[29] We use the OLS specification to examine the non-linear impact as we have been unsuccessful in finding an instrument for the quadratic of the proportion of pupils who are non-white that satisfies weak instrument testing.

4.5.4 Heterogeneity

On average, increasing the proportion of ethnic minority peers has a positive impact on the perceived likelihood at age 14 of attending university for white pupils.

The next stage of the analysis explores the extent to which the impact of peers' ethnicity differs across observable characteristics. To do this we conduct a series of subgroup analyses based on individual and family characteristics.[30] Due to small sample sizes, the sub-sample analysis provides suggestive evidence of how impact varies across observable characteristics.

We start by examining the impact across gender. Females are much more likely to go to university than males and have been for many years. The higher education participation level for females is 56.6%, compared to 44.1% for males (Hewitt 2020). Assuming that the reported likelihood of attending university is a good predictor of actual university attendance, if we find a more positive impact on male pupils then increasing the exposure to non-white pupils could aid in reducing the gender differences in university participation.

Table 4.6: The impact of ethnic minority peers on the perceived likelihood of attending university for white pupils, by gender. (Estimated by a just-identified IV model)

	(Female)	(Male)
Non-white	0.4491**	0.4618**
	(0.2128)	(0.2157)
N	1831	1776

*Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table 4.6 shows a positive impact of ethnic minority peers for both males and females, of very similar magnitudes. Diette & Oyelere (2014) and Legewie & DiPrete (2012), who both examine the impact of peers on educational achievement, find males to be most impacted. Dickerson et al. (2018), who focus on aspirations, shows that female and males' aspirations at age 14 are both impacted by their peers.

[30] We estimate all subgroup analysis using the just-identified IV model. First stage F statistics remain above 10. The average F stat across the sub-group analysis is 48.1.

We also consider the effects for white pupils who are more exposed to ethnic minority peers, for those who are economically disadvantaged, and those who performed poorly in tests at age 11. More generally, Table 4.7 shows sub-group analysis across the income distribution whilst Table 4.8 presents the impact for individuals whose parents, at best, have school level qualifications or lower compared to parents with higher levels of qualifications, Table 4.9 estimates the impact across the ability distribution.[31]

Table 4.7 shows that the estimated impact of non-white peers on white pupils' university likelihood decreases across the income distribution. Individuals with household income in the bottom 20% of the distribution are impacted more by non-white peers than individuals in the top 20% of the income distribution.

The estimated impact for pupils at the bottom of the income distribution is three times larger than the estimated impact for those in the top of the income distribution. This difference just fails to achieve statistical significance at conventional levels. We only find a significant impact for those in the middle 60% of the income distribution. We interpret these coefficients with caution as we have small sample sizes and large standard errors.

Similarly, Table 4.8 shows that the estimated impact is larger for individuals who have parents with lower level of qualifications.

As discussed in Section 4.2, the literature provides varying results. It is the same for analysing the heterogenous impacts. Hoxby (2000) and Gould et al. (2009) argue that disadvantaged pupils are more responsive to changes in school resources. On the other hand, Geay et al. (2013) provides evidence to show that disadvantaged pupils are less affected. The most recent study in the literature that focuses on aspirations, Gagete-Miranda (2022), shows that peer impacts are homogenous across individual and family characteristics.

[31] To examine the impact across the income and ability distribution we generate three dummy variables relating to the bottom 20%, top 20% and middle 60% of the distribution.

Table 4.7: The impact of ethnic minority peers on the perceived likelihood of attending university of white pupils, by household income (Estimated by a just-identified IV model)

	(Top)	(Middle)	(Bottom)
Non-white	0.2013	0.5596**	0.6059
	(0.1828)	(0.2640)	(0.3739)
N	722	2162	723

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: The impact of ethnic minority peers on university aspirations of white pupils, by parental education. (Estimated by a just-identified IV model.)

	(School level qualification)	(Higher than school qualification)
Non-white	0.5276**	0.3943*
	(0.2095)	(0.2212)
N	1401	2206

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Similarly, we can consider how the impact of non-white peers on white pupils' reported university likelihood differs across the ability distribution. Wiseman et al. (2017) shows that prior educational attainment is the main predictor of university participation and accounts for much of the variation in participation. The higher the number of GCSEs attained, the more likely pupils are to attend university. We therefore assess the impact of peers' ethnicity across the ability distribution. Table 4.9 shows that as KS2 test scores increase, the impact of peers on university likelihood decreases. The estimated effect for individuals at the lower end of the ability distribution is twice as large as for those at the upper end. We also observe a significant effect for those in the middle of the distribution.

Table 4.9: The impact of ethnic minority peers on the perceived likelihood of attending university for white pupils, by ability. (Estimated by a just-identified IV model)

	(Top)	(Middle)	(Bottom)
Non-white	0.3238	0.3938*	0.6679
	(0.2667)	(0.2162)	(0.4597)
N	830	1930	847

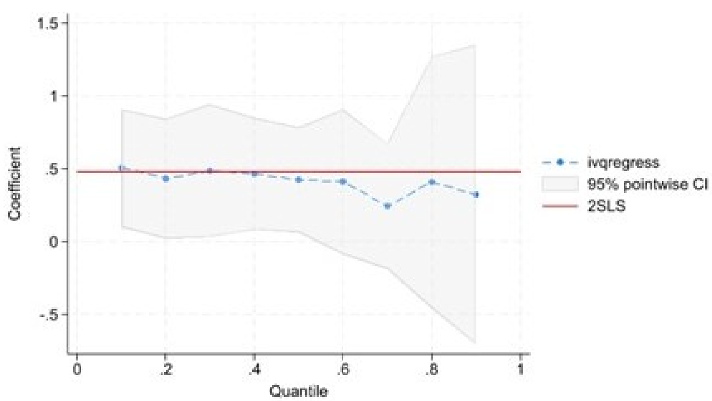
Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5.5 Aspiration distribution

We additionally examine the impact of ethnic minority peers across the distribution of reported university likelihood. If exposure to ethnic minority peers is only having an impact on individuals who already have high aspirations then the policy response would be different to if the impact was focused at the bottom of the distribution.

Figure 4.3 presents results from an instrumental variable quantile regression. Figure 4.3 shows us that the impact of ethnic minority peers is close to the 2SLS estimate across the aspiration distribution but only significant for the bottom half of the aspiration distribution. However, we should be cautious in our interpretation as the confidence intervals increase as we move up the distribution.

Figure 4.3: The impact of ethnic minority peers on the distribution of university aspirations of white pupils. (Estimated by a just-identified quantile IV model.)



Notes: Reported university likelihood is separated in to 9 quantiles. The estimated coefficient provides the impact on ethnic minority peers on individual reported likelihood of attending university from a IV quantile version of equation 4.3.

4.5.6 Ethnic groups

Non-white pupils represent a heterogeneous group. To determine whether the peer effect differs across ethnic groups we disaggregate them into the 19 ethnic groups used by the census. Whilst non-white pupils on average have higher perceived likelihood of attending university than white pupils, there is variation within non-white pupils. We may expect larger peer effects from ethnic groups with the highest reported university likelihood. We focus our analysis on the share of Indian, Pakistani and African pupils within schools as they are the largest minority groups in England.[32] We groups all other ethnic groups in to one category.[33] We estimate separate equations for each group.

Whilst our baseline results suggest an impact of 0.46, disaggregating by ethnicity produces much larger estimates. Being exposed to Indian, Pakistani and African peers increases university likelihood by between 2 and 1.5 percentage points. These findings fit with the differences in university aspirations of these ethnic groups. The average effect observed earlier therefore seems to be pushed down by the impact of all other ethnicities. As Indian, Pakistani and African pupils make up the majority of non-white peers, we suggest that the impact for some white pupils could be larger than estimated in our baseline.

Table 4.10: The impact of different groups of ethnic minority peers on the perceived likelihood of attending university for white pupils. (Estimated by a just-identified IV model)

	(Indian)	(Pakistani)	(Black African)	(Other)
Ethnic groups	2.0006**	1.4945 **	1.9291**	0.235
	(0.8295)	(0.5920)	(0.7291)	(0.4201)
N	2870	2996	2858	3033

*Notes: See Table 4.5 'Ethnic groups' is the percentage of peers in the ethnic group given by the respective column headings. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

4.5.7 Fractionalisation index

In our baseline estimation, we measure the exposure to ethnic minorities as the proportion of pupils in the school who identify as having a non-white ethnicity. While this is a simple

[32] 16.1% of the population is non-white. 3.1% of the population is Indian, 2.7% are Pakistani and 2.5% are Black African (Office of National Statisitcs 2023B).

[33] The other group include Bangladeshi, other Asian ethnicities, Black Caribbean, other black ethnicities, White and Black African, White and Asian, other mixed ethnicities and other ethnicities.

measure, it potentially lacks the depth needed to capture the full picture of ethnic diversity. We therefore use the fractionalisation index to provided a more comprehensive measure of ethnic diversity as it considers the number and size of different ethnic groups.

The fractionalization index quantifies the level of diversity within a school, with values ranging from 0 to 1. A score of 0 represents complete homogeneity, while 1 indicates maximum diversity. The index is determined by calculating the likelihood that two randomly chosen individuals from the school are from different ethnic groups.[34]

We interpret the estimates provided in Table 4.11 as a 0.1 increase in the fractionalisation index would be associated with a 4.06 percentage point increases in university likelihood.

Table 4.11: Just-identified IV model on the impact of diversity

	(University likelihood)
Diversity	40.6828***
	(14.6307)
N	3607

Notes: The dependent variable is the reported likelihood at age 14 of attending university (provided as a percentage). Diversity is the fractionalisation index which ranges from 0 to 1, where 0 indicates complete homogeneity and 1 signifies maximum diversity. Controls include gender, a proxy for ability, month of birth, main parent's age at birth, number of siblings, main parent's education and labour market status, household equalised income, the local deprivation index, the number of pupils enrolled in the school, and a proxy for school quality. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[34]The fractionalisation index is calculated using the formula:

$$Diversity = 1 - \sum_{i=1}^{N} p_i^2$$

where p_i is the proportion of the population in group i, and N is the total number of groups. Ethnic groups are defined using the 19 groups in the 2021 Census.

4.5.8 Robustness Checks

Weak instrument robustness check

As discussed in Chapter 2 Section 2.5.5, instruments are deemed weak if their correlation with the endogenous regressor, after adjusting for any controls, is nearly zero. When this correlation is low, standard approximations to the distribution of IV estimators, like two-stage least squares, become unreliable. Consequently, IV estimators may be biased, and standard IV confidence intervals may not capture the true parameter value as often as expected.

The F > 10 advice has been widely adopted in practice despite many theoretical analysts arguing for higher levels of acceptance. Andrews et al. (2019) compared the F statistics of 17 papers published in the American Economic Review from 2014 to 2018 that use instrumental variables. They show that the majority of studies present F statistics between 10-15 with some even below 10.

The first stage F statistic in this research is 59.40, higher than in many studies discussed in Andrews et al. (2019). We can reject the hypothesis that the relative (vs. OLS) asymptotic bias could be greater than 5%.

Lee et al. (2022) expands on the existing literature by introducing a new method for valid t-ratio inference in IV models.

They propose the tF critical value function, which adjusts standard errors in a way that smoothly responds to the first-stage F-statistic. Numerous studies, both theoretical and numerical, have shown that the IV t-ratio test often leads to substantial over-rejection and produces confidence intervals that under-cover the true parameter when instruments are weak. To address this, researchers typically use the first-stage F-statistic as a pre-test for instrument weakness.

Rather than relying on fixed thresholds for pre-testing, Lee et al. (2022) suggests a method that adjusts t-ratio inference in a continuous manner based on the first-stage F-statistic. This involves applying an adjustment factor to the 2SLS standard errors, which is determined by the value of the first-stage F-statistic.

With an F statistic of 59.40, we calculate our adjustment factor of 1.07. Multiplying the standard error by the adjustment factors increases the standard error slightly but the estimated coefficient remains significant at a 1% significance level.

Alternative instruments

Adding more instruments increases the asymptotic efficiency of the 2SLS estimator. The finite sample bias in 2SLS can however, get much worse by adding too many instruments (Bound et al. 1995).

For reasons explained in Section 4.4, we use the proportion of nurses in the local government district in 1951 to instrument the proportion of non-white pupils in a school in 2014 and estimate a just-identified model to estimate our baseline results.

As a robustness check we explore some of the other job categories in the 1951 census as potential additional instrumental variables. As mentioned in Section 4.4, in the post-war period, Britain was facing labour shortages in a number of occupations. In a similar story to the NHS requirement drive, in February of 1956, recruitment for transport workers began in the Caribbean. The government of Barbados even lent recruits the fare to Britain. Over the years, direct recruitments extended to other commonwealth countries.

Table 4.12 presents the IV estimates of an over-identified model using both the proportion of employed individuals in the Local Government District that worked as nurses or transport workers in 1951. We find very similar estimates to our baseline model.[35]

Whilst NHS and transport workers were the two professions specifically recruited from commonwealth countries by the government, many other industries saw increases in immigrant workers. The Empire Windrush, which arrived in Tilbury in 1948, was one of many ships that carried migrants from the Caribbean to Britain in the post-war period. The passenger list is an invaluable record of the passengers on board the ship. This includes names, ages, occupations and onward addresses. As an additional robustness check, we use the passenger list along with reports published by the National Assistance Board on the dispersal of Windrush arrivals to identify occupations/industries most commonly carried out by the passengers on the ship. These include nurses, transport workers, mechanics, textile workers, typists, tailors, defence, domestic servants, machinists and carpenters. Table 4.13 presents the IV estimates from an over-identified model, when we use these occupations as an instrument for the proportion of non-white peers in 2014. The estimated impact of non-white peers on white pupils' university likelihood remains positive and significant though reduced from 0.45 to 0.23[36]. We argue that the wide range of occupations does not work as well as an instrument compared to nurses alone due to geographical concentrations of some of the occupations. For example, the motor industry was concentrated in Birmingham. These occupations therefore do not capture the spread of ethnic minorities across the country.

Due to the robustness of the baseline findings to the inclusion of more instruments we carry out the rest of the robustness checks using the just-identified model.

[35]The F stat is 30.78 and over identification tests are satisfied.

[36]The F stat is 21.88 and over identification tests are satisfied.

Table 4.12: Over identified model using nurses and transport worker as instruments

	(University likelihood)
Non-white	0.4516***
	(0.1471)
N	3607

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.13: Over identified IV model using occupations identified on Windrush passenger lists as instruments

	(University likelihood)
Non-white	0.2257***
	(0.0732)
N	3607

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The definition of white pupils

For our baseline results, we restrict the sample to pupils with a white ethnicity. This allows us to examine the impact of the minority group on the majority. Here we check the sensitivity of the baseline results to the definition of white pupils. In the main estimation, we define white pupils as any pupil who self-identifies as belonging to the racial category commonly referred to as "White". This definition is broad and includes pupils from many geographical origins including Europe, the United States, Canada, Australia and New Zealand. We may expect the impact of non-white peers to be different for white British pupils and white people from other nations due to non-British pupils being classed as immigrants.

As a robustness check we restrict the sample to white British pupils. Table 4.14 presents the IV estimates of the slightly smaller sample. The results are very robust, with only a very small change in the magnitude of the estimate. The F statistic also remains high at 52.55.

Table 4.14: Just-identified model using only white British pupils

	(University likelihood)
Non-white	0.4599***
	(0.1637)
N	3536

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Pupils with mixed ethnicity

As discussed in Section 4.3, to calculate the percentage of pupils within a school that have a non-white ethnicity, we categorise ethnicities in to 5 groups (White, Asian, Black, Mixed and Other). The mixed ethnicity category is made-up of individuals who identify as "White and Black African" or "White and Asian". For our baseline estimates, we include this group in the non-white category. As a robustness check we remove the mixed ethnicity category from the calculations of the non-white peer group. Despite individuals who have a mixed ethnicity reporting higher likelihoods of attending university than white individuals, if these pupils are driving the results if would be difficult to attribute the impact to exposure to different ethnicities.

Table 4.15 presents the IV estimates where non-white pupils are individuals who identify as having an Asian, Black or Other ethnicity. The results are very robust, with only a very small increase in the magnitude of the estimate.[37]

Table 4.15: Just-identified model removing mixed ethnicity from measures of non-white peers

	(University likelihood)
Non-white	0.4736***
	(0.1566)
N	3607

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[37]The F statistic increases to 68.23.

Friends

In our main estimation we make the assumption that a pupil's reference group is all pupils within the school. We also assume that all pupils are equally connected and have the same influence on each other. As discussed in Section 4.2, there is evidence to suggest that homophily plays an important role in friendship formation and it is these friendship groups that have the largest impact.

At age 14, cohort members are asked how many of their friends are from the same ethnic group as them. The response is recorded in a categorical variable including, all of them, most of them, some of them and none of them. As a robustness check we drop cohort members who report that all of their friends are from the same ethnic group. This means that every pupil left in the sample says they have some close connection to a peer who is of a different ethnicity.

Table 4.16 presents the IV estimates for this smaller sample. The results are robust to the change in the sample.[38]

Table 4.16: Just-identified model on sample of pupils who report having close fiends from a different ethnic group

	(University likelihood)
Non-white	0.4078**
	(0.1590)
N	2884

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Omitted variables

All specifications control for individual, family, school and local characteristics. Despite controlling for a good amount of selection into schools, there is however still potential for

[38] The F statistics is 57.71

selection on unobservables. As school attended is normally a decision made by the parents of a child, the ethnic composition of that school may have impacted the parents' decision due to beliefs held on ethnic minorities. Dustmann & Preston (2001), using data from England, finds that individuals with greater hostility toward ethnic minorities tend to avoid settling in neighborhoods with a high concentration of ethnic minorities.

We use two measures of parental attitudes. Firstly, when the cohort member is 9 months old, years prior to them starting school, the parent is asked how they would feel about their child attending a 50/50 mixed raced school. The responses are recorded in a categorical variable ranging from strongly agree to strongly disagree. We use this categorical variable to create three dummy variables, agree, disagree and neither agree nor disagree. We include agree and disagree in the model as a proxy for parental beliefs. Secondly, when the child is 14, parents are asked how likely they think it is that the child will attend university. This is reported as a categorical variable ranging from very unlikely to very likely.

Table 4.17 presents the IV estimates with parental preferences on mixed raced schools whilst Table 4.18 presents the IV estimates including parental aspirations. The results are very robust to the control for beliefs with very little change in the estimated coefficient.[39] There is a slight increase in the estimated impact after controlling for parental aspirations.

Table 4.17: Just-identified IV model including parental preferences on mixed raced schools

	(University likelihood)
Non-white	0.4734***
	(0.1545)
N	3607

Notes:See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[39]The F statistics is 57.25.

Table 4.18: Just-identified IV model including parental aspirations.

	(University likelihood)
Non-white	0.7378*
	(0.4146)
N	3369

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Another potential omitted variable is whether the ethnic minorities are first language English. The language skills of peers is likely to influence the impact they have. Additionally, peers' language skills can also influence a pupil's exposure to ethnic minorities by shaping educational experiences and school policies. We therefore additionally control for the percentage of pupils within the school for whom English is their second language.

Table 4.19 presents the IV estimates with this additional control. The coefficient is slightly larger. We might expect this as an increase in English second language peers is likely to be positively correlated with exposure to ethnic minorities and negatively correlated with individual aspirations if language differences lead to less interaction.

Table 4.19: Just-identified IV model including percentage of pupils within the school who have English as a second language.

	(University likelihood)
Non-white	1.1489**
	(0.4935)
N	3597

Notes: See Table 4.5 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Coefficient of proportionality

The IV estimates presented in Table 4.4 are larger than the OLS estimates presented in Table 4.3. As we expect that any bias induced by the endogeneity of the ethnic composition of the school likely results in underestimation of the positive effects of ethnic minority peers on white pupils university aspirations, that is what we would have expected.

Large differences between the sizes of OLS and IV coefficients can raise concerns about the validity of the instrument. However, as discussed in Chapter 2, Ciacci (2021) argues that simply comparing the relative sizes of OLS and IV estimates is insufficient to determine whether the IV estimate accurately reflects the true effect. They recommend calculating the coefficient of proportionality using Oster (2019) bounds to better compare these estimates. This coefficient indicates how strong the selection on unobservables must be relative to the selection on observables to justify the IV estimate using the OLS model. Ciacci (2021) suggests that low values of the coefficient of proportionality provide evidence that IV estimates are not excessively large compared to OLS estimates.

Table 4.20 presents the coefficient estimated by Oster (2019) methodology setting a negative sign of the coefficient of the coefficient of proportionality gamma and Rmax=1.

We interpret this as, if the selection on unobservables is at the largest 2.15% larger than selection on observables it is enough for the true treatment effect to have the size of the IV estimates.

Table 4.20: Coefficient of proportionality

	(University likelihood)
γ	-2.15

Notes: The output of the OLS regression is displayed in Table 4.3. The output of the IV regression is displayed in Table 4.4.

4.6 Conclusion

This research provides the first analysis of the relationship between ethnic minority peers and white pupils' perceived likelihood of attending university. We shed light on the causal relationship by creating a unique instrumental variable. We exploit the government backed recruitment campaign for nurses from the Commonwealth which began in 1949 and use historical occupation data to identify areas of England that had nurse shortages.

Our IV results suggest that the relationship between the proportion of ethnic minorities in the school and university likelihood is likely causal and the more naïve OLS estimates likely underestimate the causal effect. This underestimation fits with the evidence provided by Dustmann & Preston (2001), who found that ethnic minorities are more likely to settle in areas with a lower-than-average level of education and high ability individuals choose not to live in areas with many ethnic minorities.

We find that increasing the proportion of ethnic minorities in the school by 1 percentage point leads to an 0.46 percentage point increase in white pupils' reported university likelihood at age 14. We believe age 14 is an important time to examine the impact on reported university likelihood as at this age, it is the first time that pupils get a choice over what subjects to study when picking their GCSEs. Pupils tend to see this choice as a first step in determining their future educational pathway.

These findings are robust across different samples of individuals and multiple measures of ethnic diversity. We also compare estimates when using different occupations as instrumental variables and find consistent estimates. The robustness checks consistently suggest that the existence of weak instruments or omitted variable bias is rather unlikely.

It is difficult to make comparisons with the literature due to the small number of studies that examine the impact of peer ethnicity. Focusing on the literature that also uses instrumental variables, d'Este & Einiö (2021) find negative impacts of Asian students on non-Asian pupils' maths and English Language scores whilst Geay et al. (2013) estimate small and insignificant coefficients for the impact of non-native speakers. The fact that IV studies produce LATEs rather than average treatment effects means they are less readily comparable to estimates from other work. This could explain the differing findings. Focusing on the literature which looks at aspirations, we can draw more comparisons. Dickerson et al. (2018) and

Gagete-Miranda (2022) both find that exposure to peers with high aspirations leads to large, positive, and significant peer effects on individuals' aspirations. Bringing these two strands of the literature together we find that increasing exposure to ethnic minority peers, a group who have high aspirations levels, increases white pupils' perceived likelihood of attending university.

This research has useful policy implications. The effects of immigration on a range of outcomes remain highly debated and controversial across the world. The UK is an interesting case study for this analysis as in the 2022 Programme for International Student Assessment (PISA) the UK was the only country in Europe where second-generation immigrants outperform non-immigrant students in maths (Ingram et al. 2023). Data on net migration from 2022 shows that the increase was driven by individuals arriving from non-EU countries including India, Nigeria and Pakistan, highlighting the importance of understanding the impacts of ethnic diversity (Office of National Statisitcs 2023A). We provide the first positive evidence on the impact of ethnic minorities on white pupils' perceived likelihood of attending university.

Secondly, results from our sub-group analysis provide suggestive evidence that increasing exposure to ethnic minorities could aid in increasing university participation for under-represented groups such as males and disadvantaged students. We find homogeneous impacts across gender. In regard to household income, pupils at the bottom of the income distribution (lowest 20%) have an estimated impact three times larger than the pupils at the top of the income distribution (highest 20%).

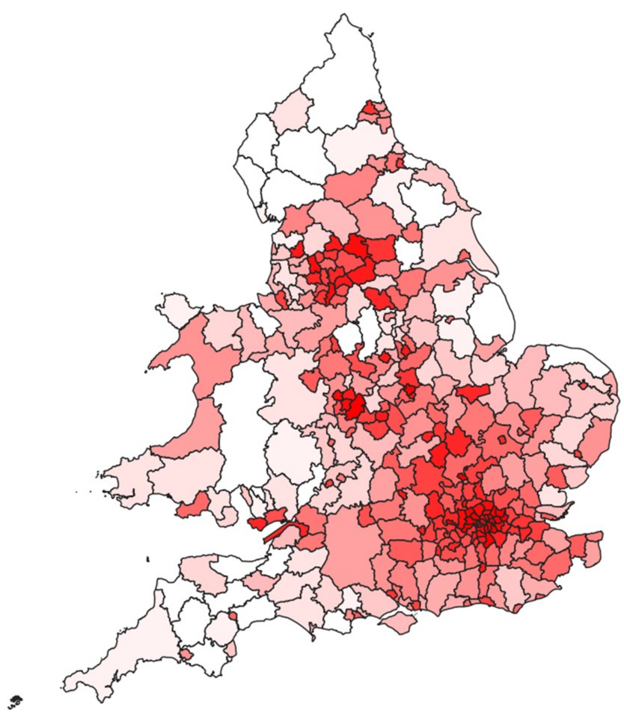
This research has been able to determine a causal relationship between ethnic minority peers and white pupils' perceived likelihood of attending university, whereby we estimate positive, significant, and robust estimates and determine the level of heterogeneity across white pupils. However, this research is limited by the sample size. Ethnic minorities are a heterogenous group, and it would be beneficial, in future research, for this analysis to go further. It is worthwhile to explore the characteristics of the ethnic minority peers and determine how this affects the estimated impacts. Furthermore, an important next step for this research is to determine the impact of white pupils on ethnic minorities. The literature to date has focused on the impacts of the minority group on the majority. To determine the overall impact, we need to understand the impact of the majority on the minority as well.

This research has contributed to the literature by investigating the causal relationship between peers' ethnicity and perceived likelihood of attending university, an area that has not been explored before. We estimate positive effects which we argue suggests that ethnic diversity in the classroom will be beneficial for all white pupils' university likelihood, and which may have the potential to increase university participation for under-represented groups such as males and specifically individuals from disadvantaged backgrounds.

Appendix C

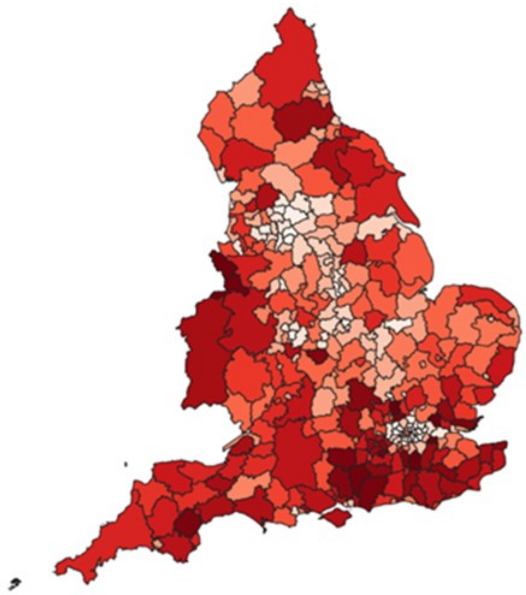
Appendix

Figure C.1: Ethnic minority density in 2011 by local authority.



Notes: The map above shows the proportion of ethnic minorities in each local authority. The darker the colour the more ethnic minorities in that area.

Figure C.2: Proportion of nurses in 1951 by local authority.



Notes: The map above shows the proportion of nurses in each local authority in 1951. The darker the colour the more nurses in that area. We assume that areas with fewer nurses in 1951 (lighter colour) have more ethnic minorities today (darker colour).

Table C.1: Average reported likelihood of attending university at age 14.

Ethnicity	Probability of attending university (age 14)
White and Black Caribbean	67.26 (30.38)
White and Black African	77.24 (24.50)
White and Asian	74.79 (25.93)
Any other mixed background	84.21 (18.29)
Indian	84.68 (20.68)
Pakistani	77.54 (24.02)
Bangladeshi	77.66 (22.77)
Any other Asian background	87.94 (17.86)
Caribbean	74.96 (23.97)
African	88.61 (15.68)
Any other Black background	82.89 (18.30)
Any other background	80.94 (22.53)

Table C.2: Levels of parents' education

Level	Qualifications
L1	Foundation Diploma GCSE(grades D–G) Scottish National level below and equal to level 4 NVQ Level 1 City and Guilds foundation-part 1 GNVQ foundation level BTEC first certification RSA level 1 Entry level qualification Level 1 Award, Basic Skill qualification Key Skill qualification YT/YTP Certificate
L2	Higher diploma O-level - GCSE (grades A*–C) Scottish National level 5 intermediate Welsh Baccalaureate NVQ Level 2 City and Guilds Craft- part 2 GNVQ intermediate BTEC level 2 RSA level 2 Level 2 Diploma/Certificate Advanced diploma A-level International Baccalaureate SCE higher Scottish Baccalaureate Advanced Welsh Baccalaureate NVQ Level 3 City and Guilds Advanced Craft GNVQ Advanced BTEC National RSA level 3 Level 3 Award/Certificate

Level	Qualifications
L3	Certificate of Higher Education NVQ Level 4 Higher National Certificate (HNC) BTEC Professional award certificate and diploma level 4 RSA level 4 Level 4 Award/ Certificate/Diploma Nursing Diploma of Higher Education Foundation degree Teaching foundation stage/ primary education/secondary education/further education Higher National Diploma (HND) BTEC Professional Award Level 5 Certificate/Diploma/ Award.
L4	First Degree Foundation degree BTEC Advanced Professional award Level 6 Certificate/Diploma/Award
L5	Master's degree Integrated master's degree BTEC Advanced Professional award Postgraduate certificate/ diploma level 7 NVQ Level 5 Doctorates

Table C.3: Just identified IV model including controls for both parents' education and labour market status

	(University likelihood)
Non-white	0.4367***
	(0.1489)
N	3607

Notes: The dependent variable is the reported likelihood at age 14 of attending university (provided as a percentage). Non-white is the percentage of peers who have a non-white ethnicity. Controls include gender, a proxy for ability, month of birth, main parent's age at birth, number of siblings, both parent's education and labour market status, household equalised income, the local deprivation index, the number of pupils enrolled in the school, and a proxy for school quality. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Just-identified tobit IV model

	(University likelihood)
Non-white	0.4808***
	(0.1568)
N	3607

Notes: The dependent variable is the reported likelihood at age 14 of attending university (provided as a percentage). Non-white is the percentage of peers who have a non-white ethnicity. Controls include gender, a proxy for ability, month of birth, main parent's age at birth, number of siblings, main parent's education and labour market status, household equalised income, the local deprivation index, the number of pupils enrolled in the school, and a proxy for school quality. Standard errors are clustered at the school level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: The impact of ethnic minority peers on the reported university likelihood of white pupils (First Stage for 2SLS estimation using proportion of nurses in 1951 as the instrumental variable)

	(Non-white)
Nurse (IV)	-1.6707***
	(0.2168)
Female	0.2196
	(0.5065)
February	1.3249
	(1.3228)
March	1.2780
	(1.1391)
April	0.3801
	(1.1322)
May	1.4180
	(1.1873)
June	1.5885
	(1.1448)
July	1.8032
	(1.1839)
August	1.2096
	(1.1287)
September	0.7974
	(1.0963)
October	0.7974
	(1.1298)
November	0.3645
	(1.0646)
December	0.0896
	(1.0378)
Ability	0.3378
	(0.2158)
Household income	0.0153***
	(0.0028)
Mother L1 education	-2.4056***
	(2.7452)
Mother L2 education	-1.6771***
	(0.6088)
Mother L3 education	-1.0888
	(0.6914)
Mother L4 education	-2.0873***
	(0.8020)
Mother L5 education	1.7255
	(1.7849)
Mother other education	8.4825***
	(2.7452)

Main employed	-3.481
	(2.9601)
Main self-employed	-3.6957
	(3.0203)
Main out of labour market	-2.8002
	(2.9564)
Main employment missing	-3.5610
	(3.0618)
Siblings	1.3721***
	(0.2793)
Main birth age	0.1511***
	(0.0468)
Deprivation 1	4.0206**
	(1.8859)
Deprivation 2	-0.5057
	(1.5367)
Deprivation 3	-0.5402
	(1.6203)
Deprivation 4	-4.3401***
	(1.4795)
Deprivation 5	-1.4262
	(1.7758)
Deprivation 6	-4.5857***
	(1.5583)
Deprivation 7	-51730***
	(1.5200)
Deprivation 8	-5.5637***
	(1.7016)
Deprivation 9	-5.1902***
	(1.7016)
School ability	0.8035***
	(0.2906)
N	3607
R^2	0.14

Notes: First stage regression. We also include the number of pupils enrolled in the school as a control. This is a categorical variables split in to many groups. Coefficient for these groups can be requested from the author. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: The impact of ethnic minority peers on the reported university likelihood of white pupils(Second Stage for 2SLS estimation)

	(University likelihood)
Non-white	0.4589***
	(0.1514)
Female	7.0953***
	(0.8898)
February	-2.0176
	(2.5887)
March	0.1068
	(2.4751)
April	0.0554
	(2.4829)
May	-1.2269
	(2.4342)
June	-4.3079*
	(2.3078)
July	-1.4775
	(2.5057)
August	0.4056
	(2.3328)
September	0.2357
	(2.4319)
October	2.1623
	(2.4319)
November	2.4375
	(2.2420)
December	3.4567
	(2.1880)
Ability	1.2767***
	(0.4340)
Household income	0.0080*
	(0.0043)
Mother L1 education	-1.4621
	(1.7046)
Mother L2 education	-3.9315***
	(1.4983)
Mother L3 education	6.1086***
	(1.4708)
Mother L4 education	11.6819***
	(1.5404)
Mother L5 education	17.5521***
	(2.3674)
Mother other education	-1.4980
	(3.6096)

	(University likelihood)
Main employed	5.2528
	(4.7456)
Main self-employed	6.1765
	(4.9180)
Main out of labour market	2.1614
	(4.8228)
Main employment missing	4.2843
	(5.1304)
Siblings	-1.1899*
	(0.5542)
Main birth age	-0.0929
	(0.0918)
Deprivation 1	-3.4852
	(2.6539)
Deprivation 2	-3.3051
	(2.6539)
Deprivation 3	-4.9518*
	(2.5422)
Deprivation 4	0.1118
	(2.7928)
Deprivation 5	-4.5569*
	(2.5652)
Deprivation 6	-1.6772
	(2.7183)
Deprivation 7	-1.3772
	(2.5407)
Deprivation 8	-0.2457
	(2.6527)
Deprivation 9	-0.4663
	(2.7323)
School ability	1.90079***
	(0.3991)
N	3607
R^2	0.09

Notes: Second stage regression. We also include the number of pupils enrolled in the school as a control This is a categorical variables split in to many groups. Coefficient for these groups can be requested from the author. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 5

Conclusion

Education outcomes are influenced by numerous factors. Using England as a case study, each chapter in this thesis examines different determinants of pupils' cognitive and non-cognitive skills, as well as beliefs about future education outcomes, including childcare, absences, and peer ethnicity. Although existing research has highlighted these areas as important, there are still large gaps in the literature. This thesis adds to the literature by addressing unanswered questions, using novel datasets, and creating new methodologies.

The first empirical chapter of this thesis looks at the effect of time spent in formal childcare by age 3 on non-cognitive skills in the short and long run. Building on the literature that shows the positive impact of attending childcare, this research is the first to determine both the short and long run impact of time spent in childcare on non-cognitive skills. The second empirical study examines the causal relationship between school absences and academic achievement. By linking administrative and survey data the research has more information on background characteristics than previous studies allowing for deeper heterogeneity analysis. Finally, the last chapter determines the impact of peers' ethnicity on white pupils' perceived likelihood of attending university. Using historical data to create a unique instrumental variable, this research is the first to shed light on the causal relationship between peers' ethnicity and individual aspirations.

5.1 Summary of results

Chapter 2 focuses on the intensity of childcare, measured in hours per week, for children under the age of 3 on non-cognitive skills at ages 3-14, which are measured by the Strength and Difficulties Questionnaire (SDQ), proposed by Goodman (1997). We use data from the Millennium Cohort Study and create instrumental variables that leverages exogenous variation in both the probability that the mother works shift work and has uncertain working hours. We show that time spent in childcare at age 3 has initial impacts on non-cognitive skills which persist into the medium run. The fact that these effects do persist underlines the importance of early childhood education. Results from our sub-group analysis suggest that there is heterogeneity in the magnitude and persistence of impacts across population subgroups. Most notably, less advantaged children, particularly those with low-educated mothers and those in lower-income households, appear to initially benefit most from hours spent in childcare. This research adds to the knowledge base on the impacts of childcare by focusing on the intensity of childcare as well as the long run impacts on non-cognitive skills.

Chapter 3 provides the first study to shed light on the causal impact of absences on end of year achievement, in the context of England. Linking the NPD to Understanding Society (UKHLS)- Harmonised British Household Panel data (BHPS), the research is able to build on the literature by having information on pupils' background and family characteristics. The research shows that on average missing one additional session of school reduces an individual's end of year attainment in Maths/English by 0.3% of a standard deviation. In addition, the research finds approximately linear effects of pupil absences on end of year achievement in Maths and English with no discontinuity at the persistently absent threshold. We find no evidence of heterogeneous impacts of absences across gender or ethnicity, but the findings do suggest that reducing the absences of those pupils at the bottom of the ability distribution could aid in reducing the performance gaps. These findings are comparable to the literature, which manly focuses of the US, allowing for greater confidence in the generalisability.

The final empirical study, presented in Chapter 4, examines the influence of ethnic minority peers on the perceived likelihood of attending university for white pupils. We use the Millennium Cohort Study to determine the reported university likelihood and the NPD to calculate the proportion of pupils within the school who report their ethnicity as non-white. We link this to historical occupation data from the 1951 census to create a unique instrumental variable based on the proportion of nurse in each area in 1951. Our two-stage-least-squares estimation suggest that the relationship between the proportion of ethnic minorities in the school and university likelihood is likely causal and the more naïve OLS estimates likely underestimate the causal effect. We find that increasing the proportion of ethnic minorities in the school by 1 percentage point leads to an 0.46 percentage point increase in white pupils' reported university likelihood at age 14. Results from our sub-group analysis provide suggestive evidence that increasing exposure to ethnic minorities could aid in increasing university participation for under-represented groups such as males and disadvantaged students. To avoid the generalisation of ethnic minorities, we split the sample by ethnic groups to unpick how the peer effect varies by exposure to different ethnic groups. Here we find that white pupils are influenced more by the largest ethnic groups in England, Indian, Pakistani, and Black African. This research not only provides a methodological contribution but also the first analysis of the relationship between ethnic minority peers and white pupils' perceived likelihood of attending university.

Overall, this thesis highlights the significance of early and sustained educational engagement across various dimensions. Together, these findings have policy implications, discussed below, that create more equitable and effective educational systems.

5.2 Policy implication and future research

This thesis uses novel datasets and robust quantitative methods to inform policymakers about the determinants of cognitive and non-cognitive skills and beliefs about future education outcomes. The aim is to provide evidence to develop the English educational system, improve the outcomes of disadvantaged pupils, and reduce inequality in education outcomes.

There are important policy implications to be drawn from this thesis. Firstly, Chapter 2, consistent with the broader literature on childcare enrolment, estimates positive effects of childcare. There is a current interest in the optimal number of hours in childcare in England due to the upcoming changes in childcare subsidies. The subsided hours in childcare for working parents are increasing from 30 hours for 3- and 4-year-olds to cover all under 5-year-olds by 2025. Whilst this policy was driven by the impact on female labour supply, our findings provide evidence to show that the policy change also has benefits for the child.

Additionally, results from our sub-group analysis suggest that there is heterogeneity in the magnitude and persistence of impacts across population subgroups. Most notably, less advantaged children—particularly those with low-educated mothers and those in lower-income households appear to initially benefit most from hours spent in childcare. These findings are consistent with a large proportion of the literature which focuses on childcare for disadvantaged children. These findings support the argument that facilitating access to more time in childcare for disadvantaged children may hold potential for decreasing early socioeconomic disparities in child development.

In future research, it would be worthwhile to explore the impact across different care settings. While we have shown that increased hours are beneficial, ensuring that these additional hours are spent in high-quality environments could be crucial for maximising positive outcomes. It would also be interesting to investigate some of the mechanisms involved in the impact. Whilst we hypothesise that the group setting of formal childcare, coupled with professional carers could be a potential mechanism, we do not formally test this due to sample size.

The results presented in Chapter 3 identify a negative and significant impact of absences on academic achievement. Consistent with the literature, we find this relationship to be linear and find no evidence of a discontinuity at the persistently absent threshold. This linearity suggests that the persistently absent indicator is arbitrary and may miss a large proportion of absences that are having a negative impact on achievement.

Also, in the 2022 School's White Paper (Department for Education 2022C), the introduction of a minimum expectation on the length of the school week of 32.5 hours was announced. It is not currently possible to analyse the impact of extending the school week in England as The Department for Education does not record schools' opening hours. However, the ability to target specific pupils who would benefit the most from reduced absences suggests that focused initiative might be more effective than extending the school week for everyone. Additionally, this could also help narrow current achievement gaps. We provide evidence to suggest that reducing the absences of those pupils at the bottom of the ability distribution could aid in reducing the performance gaps. We find that pupils at the lower end of the ability distribution miss 11 days more than those at the higher end. We also find heterogeneous impacts across the ability distribution, particularly for Maths. Reducing absences of those at the lower end of the ability distribution relative to those at the higher end could increase the average achievement of low achievers by 3.3% of a standard deviation. Whilst this alone would not correct the performance gaps, these are nontrivial reductions.

In future research, it is worthwhile to explore why pupils are absent. Within the sample of analysis only 4.6% reported the reason for the absence. Over time this information is becoming for readily available and with additional data collection from places like the FFT attendance tracker, this will allow for the reason of absence to be examined along with other questions regarding the length the absence. Understanding the reason for absence will allow for more targeted interventions. The Labour Government has announced numerous initiatives from breakfast clubs to mental health advisors, and understanding the impact of these on absences is crucial for determining their effectiveness.

Chapter 4 provides the first positive evidence on the impact of ethnic minorities on white pupils' perceived likelihood of attending university. The impact of immigration on various outcomes continues to be a highly debated and controversial topic globally.

Gorard et al. (2012), who provide a literature review on the relationship between aspirations and educational attainment and participation, describe evidence of a positive association between aspirations and university participation. Results from our sub-group analysis could therefore provide suggestive evidence that increasing exposure to ethnic minorities could aid in increasing university participation for under-represented groups such as males and disadvantaged students. We find homogeneous impacts across gender. Regarding household income, pupils at the bottom of the income distribution (lowest 20%) have an estimated impact three times larger than the pupils at the top of the income distribution (highest 20%).

Chapter 4 is able to shed light on the causal relationship between ethnic minority peers and white pupils' perceived likelihood of attending university, whereby we estimate positive, significant, and robust estimates and determine the level of heterogeneity across white pupils. However, this research is limited by the sample size. Ethnic minorities are a heterogenous group, and it would be beneficial, in future research, for this analysis to go further. It is worthwhile to explore the characteristics of the ethnic minority peers and determine how this affects the estimated impacts. Furthermore, an important next step for this research is to determine the impact of white pupils on ethnic minorities. The literature to date has focused on the impacts of the minority group on the majority. To determine the overall impact, we need to understand the impact of the majority on the minority as well. In addition, exploring the impact on university participation would be the next step. Sweep 7 of the MCS is due for release in late 2025 and would allow for this additional outcome variable. As data availability of administrative datasets is increasing, it allows for larger sample sizes to be examined.

In summary, this thesis contributes the various strands of the Economics of Education literature as set out in each individual chapter. It builds upon and develops existing academic research on the determinants of cognitive and non-cognitive skills and beliefs about future education outcomes of pupils in England. In its entirety, this thesis highlights the importance of robust methodology to shed light on the causal effects of education determinants. By using well established quasi-experimental techniques alongside the creation of unique instrumental variables, we can estimate more accurately the impact of various determinants. A common message coming out of these three empirical chapters is the larger effects for more disadvantaged pupils. Reducing inequality within education is proving difficult as despite large policy attention, many education inequalities have persisted over time. This thesis highlights just a few of the wide-ranging factors which could be influencing this inequality.

Additionally, this thesis highlights the potential long-term benefits of early investment. Chapter 2 highlights the medium run impacts of time spent in childcare at a young age, and Chapter 4 shows that the ethnic composition of a school has positive effects of beliefs about future university attendance which are likely correlated with actual attendance. Both effects were found to be larger for disadvantaged pupils, by identifying and addressing the barriers to educational success early on, we can develop a more equitable education system that promotes better outcomes for all pupils.

While there are limitations to the research presented in this thesis, it fills key gaps in the literature, creates new methodologies, provides useful policy relevant findings, and highlights several avenues for future research. This thesis not only contributes to the academic understanding of educational determinants but also offers practical insights for policymakers aiming to reduce educational disparities.

Bibliography

Agüero, J. M. & Beleche, T. (2013), 'Test-mex: Estimating the effects of school year length on student performance in mexico', *Journal of Development Economics* **103**, 353–361.

Al Baghal, T. (2016), 'Obtaining data linkage consent for children: factors influencing outcomes and potential biases', *International Journal of Social Research Methodology* **19**(6), 623–643.

Almlund, M., Duckworth, A. L., Heckman, J. & Kautz, T. (2011), *Personality Psychology and Economics*. Handbook of the Economics of Education. **4**, 1–181.

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

Anderson, T. W. & Rubin, H. (1949), 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *The Annals of Mathematical Statistics* **20**(1), 46–63.

Andrew, A., Bandiera, O., Costa-Dias, M. & Landais, C. (2021), 'Women and men at work.'. London: IFS. Accessed: July 10, 2023. `https://ifs.org.uk/inequality/wp-content/uploads/2021/12/IFS-Inequality-Review-women-and-men-at-work.pdf`.

Andrews, I., Stock, J. H. & Sun, L. (2019), 'Weak instruments in instrumental variables regression: Theory and practice', *Annual Review of Economics* **11**(1), 727–753.

Andrews, J., Robinson, D. & Hutchinson, J. (2017), 'Closing the gap? trends in educational attainment and disadvantage'. London: Education Policy Institute.

Angrist, J. D. & Lang, K. (2004), 'Does school integration generate peer effects? evidence from boston's metco program', *American Economic Review* **94**(5), 1613–1634.

Angrist, J. D. & Pischke, J.-S. (2009), 'Mostly harmless econometrics: An empiricist's companion'. Princeton university press, Princeton, New Jersey.

Angrist, J. & Kolesár, M. (2024), 'One instrument to rule them all: The bias and coverage of just-id iv', *Journal of Econometrics* **240**(2), 105398.

Apps, P., Mendolia, S. & Walker, I. (2013), 'The impact of pre-school on adolescents' outcomes: Evidence from a recent english cohort', *Economics of Education Review* **37**, 183–199.

Argys, L. M. & Rees, D. I. (2008), 'Searching for peer group effects: A test of the contagion hypothesis', *The Review of Economics and Statistics* **90**(3), 442–458.

Arulampalam, W., Naylor, R. A. & Smith, J. (2012), 'Am i missing something? the effects of absence from class on student performance', *Economics of Education Review* **31**(4), 363–375.

Aucejo, E. M. & Romano, T. F. (2016), 'Assessing the effect of school days and absences on test score performance', *Economics of Education Review* **55**, 70–87.

Babikian, C. (2021), '"partnership not prejudice": British nurses, colonial students, and the national health service, 1948–1962', *Journal of British Studies* **60**(1), 140–168.

Baker, M. (2011), 'Innis lecture: Universal early childhood interventions: what is the evidence base?', *Canadian Journal of Economics/Revue canadienne d'économique* **44**(4), 1069–1105.

Baker, M., Gruber, J. & Milligan, K. (2019), 'The long-run impacts of a universal child care program', *American Economic Journal: Economic Policy* **11**(3), 1–26.

Ballatore, R. M., Fort, M. & Ichino, A. (2018), 'Tower of babel in the classroom: immigrants and natives in italian schools', *Journal of Labor Economics* **36**(4), 885–921.

Becker, G. S. (1962), 'Investment in human capital: A theoretical analysis', *Journal of Political Economy* **70**(5, Part 2), 9–49.

Becker, G. S., Murphy, K. M. & Tamura, R. (1990), 'Human capital, fertility, and economic growth', *Journal of Political Economy* **98**(5, Part 2), S12–S37.

Belfield, C. R., Nores, M., Barnett, S. & Schweinhart, L. (2006), 'The high/scope perry preschool program: Cost–benefit analysis using data from the age-40 followup', *Journal of Human Resources* **41**(1), 162–190.

Berger, L. M., Panico, L. & Solaz, A. (2021), 'The impact of center-based childcare attendance on early child development: Evidence from the french elfe cohort', *Demography* **58**(2), 419–450.

Bernal, R. & Keane, M. P. (2010), 'Quasi-structural estimation of a model of childcare choices and child cognitive ability production', *Journal of Econometrics* **156**(1), 164–189.

Bernal, R. & Keane, M. P. (2011), 'Child care choices and children's cognitive achievement: The case of single mothers', *Journal of Labor Economics* **29**(3), 459–512.

Betts, J. R. & Fairlie, R. W. (2001), 'Explaining ethnic, racial, and immigrant differences in private school attendance', *Journal of Urban Economics* **50**(1), 26–51.

Betts, J. R. & Zau, A. (2004), 'Peer groups and academic achievement: Panel evidence from administrative data', *Unpublished manuscript* .

Bianchi, S. M. (2000), 'Maternal employment and time with children: Dramatic change or surprising continuity?', *Demography* **37**(4), 401–414.

Bifulco, R., Fletcher, J. M. & Ross, S. L. (2011), 'The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health', *American Economic Journal: Economic Policy* **3**(1), 25–53.

Biosca, O., Lenton, P. & Mosley, P. (2014), 'Where is the 'plus' in 'credit-plus'? the case of chiapas, mexico', *The Journal of Development Studies* **50**(12), 1700–1716.

Black, S. E., Devereux, P. J. & Salvanes, K. G. (2007), 'From the cradle to the labor market? the effect of birth weight on adult outcomes', *The Quarterly Journal of Economics* **122**(1), 409–439.

Blanchflower, D. G. & Oswald, A. J. (2004), 'Well-being over time in britain and the usa', *Journal of Public Economics* **88**(7-8), 1359–1386.

Blanden, J., Del Bono, E., Hansen, K. & Rabe, B. (2022), 'Quantity and quality of childcare and children's educational outcomes', *Journal of Population Economics* **35**(2), 785–828.

Blanden, J., Del Bono, E., McNally, S. & Rabe, B. (2016), 'Universal pre-school education: The case of public funding with private provision', *The Economic Journal* **126**(592), 682–723.

Booij, A. S., Leuven, E. & Oosterbeek, H. (2017), 'Ability peer effects in university: Evidence from a randomized experiment', *The Review of Economic Studies* **84**(2), 547–578.

Borghans, L., Duckworth, A. L., Heckman, J. J. & Ter Weel, B. (2008), 'The economics and psychology of personality traits', *Journal of Human Resources* **43**(4), 972–1059.

Borgonovi, F., Choi, Á. & Paccagnella, M. (2021), 'The evolution of gender gaps in numeracy and literacy between childhood and young adulthood', *Economics of Education Review* **82**, 102119.

Bossavie, L. (2020), 'The effect of immigration on natives' school performance: Does length of stay in the host country matter?', *Journal of Human Resources* **55**(2), 733–766.

Bound, J., Jaeger, D. A. & Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American statistical association* **90**(430), 443–450.

Bowlby, J. et al. (1951), *Maternal care and mental health*, Vol. 2, World Health Organization Geneva.

Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P. & Rogers, B. W. (2012), 'Homophily and long-run integration in social networks', *Journal of Economic Theory* **147**(5), 1754–1786.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**(1), 41–55.

Britton, J., van der Erve, L., Waltmann, B. & Xu, X. (2021), 'London calling? higher education, geographical mobility and early-career earnings', *Institute for Fiscal Studies Report* . Accessed December 2023. `https://ifs.org.uk/publications/london-calling-higher-education-geographical-mobility-and-early-career-earnings`.

Burger, K. (2010), 'How does early childhood care and education affect cognitive development? an international review of the effects of early interventions for children from different social backgrounds', *Early Childhood Research Quarterly* **25**(2), 140–165.

Burgess, S. M. (2014), 'Understanding the success of london's schools'. Working Paper No. 14/333. Bristol: The Center for Market and Public Organisation.

Burke, M. A. & Sass, T. R. (2013), 'Classroom peer effects and student achievement', *Journal of Labor Economics* **31**(1), 51–82.

Bursztyn, L., Ederer, F., Ferman, B. & Yuchtman, N. (2014), 'Understanding peer effects in financial decisions: Evidence from a field experiment', *Econometrica* **82**(4), 1273–1301.

Caprera, G. D. (2015), 'Ready to learn: the role of childcare attendance on children's school outcomes in italy'. CEIS Working Paper No. 378.

Card, D. & Rothstein, J. (2007), 'Racial segregation and the black–white test score gap', *Journal of Public Economics* **91**(11-12), 2158–2184.

Carlsson, M., Dahl, G. B., Öckert, B. & Rooth, D.-O. (2015), 'The effect of schooling on cognitive skills', *Review of Economics and Statistics* **97**(3), 533–547.

Carneiro, P., Cattan, S. & Ridpath, N. (2024), 'The short-and medium-term impacts of sure start on educational outcomes'. London: Institute for Fiscal Studies.

Carneiro, P. M. & Heckman, J. J. (2003), 'Human capital policy'. Germany: IZA Discussion paper No. 821.

Carrell, S. E. & Hoekstra, M. L. (2010), 'Externalities in the classroom: How children exposed to domestic violence affect everyone's kids', *American Economic Journal: Applied Economics* **2**(1), 211–228.

Carrell, S. E., Hoekstra, M. & West, J. E. (2011), 'Is poor fitness contagious?: Evidence from randomly assigned friends', *Journal of Public Economics* **95**(7-8), 657–663.

Carta, F. & Rizzica, L. (2018), 'Early kindergarten, maternal labor supply and children's outcomes: evidence from italy', *Journal of Public Economics* **158**, 79–102.

Cascio, E. U. & Lewis, E. G. (2012), 'Cracks in the melting pot: immigration, school choice, and segregation', *American Economic Journal: Economic Policy* **4**(3), 91–117.

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

Catney, G., Lloyd, C. D., Ellis, M., Wright, R., Finney, N., Jivraj, S. & Manley, D. (2023), 'Ethnic diversification and neighbourhood mixing: A rapid response analysis of the 2021 census of england and wales', *The Geographical Journal* **189**(1), 63–77.

Cattan, S., Kamhöfer, D. A., Karlsson, M. & Nilsson, T. (2023), 'The long-term effects of student absence: Evidence from sweden', *The Economic Journal* **133**(650), 888–903.

Caughy, M. O., DiPietro, J. A. & Strobino, D. M. (1994), 'Day-care participation as a protective factor in the cognitive development of low-income children', *Child Development* **65**(2), 457–471.

Chareyron, S., Chung, A. & Domingues, P. (2021), 'Ethnic diversity and educational success: Evidence from france', *Research in Economics* **75**(2), 133–143.

Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014), 'Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood', *American Economic Review* **104**(9), 2633–2679.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. & Vignoles, A. (2013), 'Widening participation in higher education: analysis using linked administrative data', *Journal of the Royal Statistical Society Series A: Statistics in Society* **176**(2), 431–457.

Ciacci, R. (2021), *A Matter of Size: Comparing IV and OLS estimates*, Universidad Pontificia Comillas. Unpublished doctoral thesis.

Commission on Race and Ethnic Disparities (2021), 'The report'. Accessed June 2023. `https://www.gov.uk/government/publications/the-report-of-the-commission-on-race-and-ethnic-disparities`.

Contreras, D. & Gallardo, S. (2022), 'The effects of mass migration on the academic performance of native students. evidence from chile', *Economics of Education Review* **91**, 102314.

Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996), 'The effects of summer vacation on achievement test scores: A narrative and meta-analytic review', *Review of Educational Research* **66**(3), 227–268.

Cornelissen, T., Dustmann, C., Raute, A. & Schönberg, U. (2018), 'Who benefits from universal child care? estimating marginal returns to early child care attendance', *Journal of Political Economy* **126**(6), 2356–2409.

Crawford, C. & Greaves, E. (2015), 'Socio-economic, ethnic and gender differences in he participation'. London: Department for Business, Innovation & Skills.

Cryan, J. R., Sheehan, R., Wiechel, J. & Bandy-Hedden, I. G. (1992), 'Success outcomes of full-day kindergarten: More positive behavior and increased achievement in the years after', *Early Childhood Research Quarterly* **7**(2), 187–203.

Currarini, S., Jackson, M. O. & Pin, P. (2009), 'An economic model of friendship: Homophily, minorities, and segregation', *Econometrica* **77**(4), 1003–1045.

Currarini, S., Matheson, J. & Vega-Redondo, F. (2016), 'A simple model of homophily in social networks', *European Economic Review* **90**, 18–39.

Dahl, G. B., Løken, K. V. & Mogstad, M. (2014), 'Peer effects in program participation', *American Economic Review* **104**(7), 2049–2074.

De Bruin, W. B., Fischhoff, B., Millstein, S. G. & Halpern-Felsher, B. L. (2000), 'Verbal and numerical expressions of probability:"it's a fifty–fifty chance"', *Organizational behavior and human decision processes* **81**(1), 115–131.

De Giorgi, G., Pellizzari, M. & Redaelli, S. (2009), 'Be as careful of the company you keep as of the books you read: peer effects in education and on the labor market'. NBER Working Papers 14948. National Bureau of Economic Research.

De Giorgi, G., Pellizzari, M. & Redaelli, S. (2010), 'Identification of social interactions through partially overlapping peer groups', *American Economic Journal: Applied Economics* **2**(2), 241–275.

Dearden, L. & Sibieta, L. (2010), 'Ethnic inequalities in child outcomes.'. London: Institute of Education. Accessed June 2023, `https://cls.ucl.ac.uk/wp-content/uploads/2017/05/12_briefing_web.pdf`.

Dee, T. S. (2024), 'Higher chronic absenteeism threatens academic recovery from the covid-19 pandemic', *Proceedings of the National Academy of Sciences* **121**(3), e2312249121.

Del Boca, D., Pasqua, S. & Suardi, S. (2016), 'Child care, maternal employment, and children's school outcomes. an analysis of italian data', *European Journal of Population* **32**, 211–229.

Del Boca, D., Piazzalunga, D. & Pronzato, C. (2018), 'The role of grandparenting in early childcare and child outcomes', *Review of Economics of the Household* **16**, 477–512.

Deming, D. (2009), 'Early childhood intervention and life-cycle skill development: Evidence from head start', *American Economic Journal: Applied Economics* **1**(3), 111–134.

Department for Education (2016), 'The link between absence and attainment at ks2 and ks4. 2013/14 academic year.'. Accessed March 2022. `https://www.gov.uk/government/publications/absence-and-attainment-at-key-stages-2-and-4-2012-to-2013`.

Department for Education (2019), 'Pupil absences in schools in england academic year 2018/19.'. Accessed March 2022. `https://www.gov.uk/government/statistics/pupil-absence-in-schools-in-england-2018-to-2019`.

Department for Education (2022A), 'Attendance action alliance'. Accessed September 2022. `https://www.gov.uk/government/news/education-secretary-launches-new-attendance-alliance`.

Department for Education (2022B), 'Schools, pupils and their characteristics'. Accessed September 2022. `https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2021`.

Department for Education (2022C), 'March 2022 schools white paper'. Accessed September 2022. `https://commonslibrary.parliament.uk/research-briefings/cbp-9511/`.

Department for Education (2024), 'Schools, pupils and their characteristics'. Accessed June 2024. `https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics`.

d'Este, R. & Einiö, E. (2021), 'Beyond black and white: the impact of asian peers on scholastic achievement', *Economics of Education Review* **83**, 102129.

Dickerson, A., Maragkou, K. & McIntosh, S. (2018), 'The causal effect of secondary school peers on educational aspirations', *Centre for Vocational Education Research Discussion Paper* **17**.

Diette, T. M. & Oyelere, R. U. (2014), 'Gender and race heterogeneity: The impact of students with limited english on native students' performance', *American Economic Review* **104**(5), 412–417.

Diette, T. M. & Uwaifo Oyelere, R. (2017), 'Gender and racial differences in peer effects of limited english students: a story of language or ethnicity?', *IZA Journal of Migration* **6**, 1–18.

Ding, W. & Lehrer, S. F. (2007), 'Do peers affect student achievement in china's secondary schools?', *The Review of Economics and Statistics* **89**(2), 300–312.

Drange, N. & Havnes, T. (2019), 'Early childcare and cognitive development: Evidence from an assignment lottery', *Journal of Labor Economics* **37**(2), 581–620.

Drayton, E., Farquharson, C., Ogden, K., Sibieta, L., Tahir, I. & Waltmann, B. (2023), 'Annual report on education spending in england: 2023'. London: IFS Report.

Dustmann, C. & Preston, I. (2001), 'Attitudes to ethnic minorities, ethnic context and location decisions', *The Economic Journal* **111**(470), 353–373.

Education Endowment Foundation (2022), 'Attendance interventions rapid evidence assessment'. Accessed Septmeber 2022. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/attendance-interventions-rapid-evidence-assessment.

Elango, S., García, J. L., Heckman, J. J. & Hojman, A. (2015), Early childhood education, *in* 'Economics of Means-Tested Transfer Programs in the United States, Volume 2', University of Chicago Press, pp. 235–297.

Elek, C., Gubhaju, L., Lloyd-Johnsen, C., Eades, S. & Goldfeld, S. (2020), 'Can early childhood education programs support positive outcomes for indigenous children? a systematic review of the international literature', *Educational Research Review* **31**, 100363.

European Commission (2024), Government expenditure on education, Technical report, European Commission. Accessed June 2024. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Government_expenditure_on_education.

Eyles, A., Lillywhite, E. & Major, L. (2023), 'The rising tide of school absences in the post-pandemic era'. LSE Blog. Accessed September 2023. https://blogs.lse.ac.uk/politicsandpolicy/the-rising-tide-of-school-absences-in-the-post-pandemic-era/.

Farquharson, C., McKendrick, A., Ridpath, N. & Tahir, I. (2024), 'The state of education: What awaits the next government?'. London: IFS. Accessed July 2024. https://ifs.org.uk/publications/state-education-what-awaits-next-government#:~:text=The%20next%20government%20will%20inherit%20an%20education%20system%20that%20produces,well%20above%20the%20OECD%20average.

Farquharson, C., McNally, S. & Tahir, I. (2022), 'Education inequalities'. IFS Deaton Review of Inequalities. London: Institute for Fiscal Studies.

Feinstein, L. (2002), 'Quantitative estimates of the social benefits of learning, 2: Health (depression and obesity). wider benefits of learning research report.'. London: Centre for Research on the Wider Benefits of Learning, Institute of Education.

Feld, J. & Zölitz, U. (2017), 'Understanding peer effects: On the nature, estimation, and channels of peer effects', *Journal of Labor Economics* **35**(2), 387–428.

Felfe, C. & Lalive, R. (2018), 'Does early child care affect children's development?', *Journal of Public Economics* **159**, 33–53.

Felfe, C., Nollenberger, N. & Rodríguez-Planas, N. (2015), 'Can't buy mommy's love? universal childcare and children's long-term cognitive development', *Journal of Population Economics* **28**, 393–422.

Fernández-Reino, M. (2016), 'Immigrant optimism or anticipated discrimination? explaining the first educational transition of ethnic minorities in england', *Research in Social Stratification and Mobility* **46**, 141–156.

Figlio, D., Giuliano, P., Marchingiglio, R., Ozek, U. & Sapienza, P. (2024), 'Diversity in schools: Immigrants and the educational performance of us-born students', *Review of Economic Studies* **91**(2), 972–1006.

Figlio, D. N. & Kenny, L. W. (2007), 'Individual teacher incentives and student performance', *Journal of Public Economics* **91**(5-6), 901–914.

Figlio, D. & Özek, U. (2019), 'Unwelcome guests? the effects of refugees on the educational outcomes of incumbent students', *Journal of Labor Economics* **37**(4), 1061–1096.

Fischer, M., Karlsson, M., Nilsson, T. & Schwarz, N. (2020), 'The long-term effects of long terms–compulsory schooling reforms in sweden', *Journal of the European Economic Association* **18**(6), 2776–2823.

Fitzpatrick, M. D., Grissmer, D. & Hastedt, S. (2011), 'What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment', *Economics of Education Review* **30**(2), 269–279.

Fort, M., Ichino, A. & Zanella, G. (2020), 'Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families', *Journal of Political Economy* **128**(1), 158–205.

Fryer Jr, R. G. & Torelli, P. (2010), 'An empirical analysis of 'acting white'', *Journal of Public Economics* **94**(5-6), 380–396.

Fryer, R. G. (2013), 'Teacher incentives and student achievement: Evidence from new york city public schools', *Journal of Labor Economics* **31**(2), 373–407.

Gagete-Miranda, J. (2022), 'An aspiring friend is a friend indeed: On the mechanisms behind peer influences on human capital accumulation', *Manuscript: Bocconi University: Milan, Italy* .

Garces, E., Thomas, D. & Currie, J. (2002), 'Longer-term effects of head start', *American Economic Review* **92**(4), 999–1012.

Gaviria, A. & Raphael, S. (2001), 'School-based peer effects and juvenile behavior', *Review of Economics and Statistics* **83**(2), 257–268.

Geay, C., McNally, S. & Telhaj, S. (2013), 'Non-native speakers of english in the classroom: what are the effects on pupil performance?', *The Economic Journal* **123**(570), F281–F307.

Geoffroy, M.-C., Côté, S. M., Giguère, C.-É., Dionne, G., Zelazo, P. D., Tremblay, R. E., Boivin, M. & Séguin, J. R. (2010), 'Closing the gap in academic readiness and achievement: The role of early childcare', *Journal of Child Psychology and Psychiatry* **51**(12), 1359–1367.

Gershenson, S., Jacknowitz, A. & Brannegan, A. (2017), 'Are student absences worth the worry in us primary schools?', *Education Finance and Policy* **12**(2), 137–165.

Gibbons, S. & Telhaj, S. (2008), ‘Peers and achievement in england’s secondary schools’. London School of Economics Spatial Economics Research Centre SERCDP0001.

Goodman, A. & Sianesi, B. (2005), ‘Early education and children’s outcomes: how long do the impacts last?’, *Fiscal Studies* **26**(4), 513–548.

Goodman, J. (2014), ‘Flaking out: Student absences and snow days as disruptions of instructional time’. NBER Working Papers 20221. National Bureau of Economic Research.

Goodman, R. (1997), ‘The strengths and difficulties questionnaire: a research note’, *Journal of Child Psychology and Psychiatry* **38**(5), 581–586.

Gorard, S., See, B. H., Davies, P. et al. (2012), ‘The impact of attitudes and aspirations on educational attainment and participation’, *York: Joseph Rowntree Foundation* .

Gottfried, M. A. (2010), ‘Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach’, *American Educational Research Journal* **47**(2), 434–465.

Gottfried, M. A. & Kirksey, J. J. (2017), ‘“when” students miss school: The role of timing of absenteeism on students’ test performance’, *Educational Researcher* **46**(3), 119–130.

Gould, E. D., Lavy, V. & Daniele Paserman, M. (2009), ‘Does immigration affect the long-term educational outcomes of natives? quasi-experimental evidence’, *The Economic Journal* **119**(540), 1243–1269.

Green, C. & Iversen, J. M. V. (2022), ‘Refugees and the educational attainment of natives: Evidence from norway’, *Economics of Education Review* **88**, 102–258.

Green, F. (2024), ‘Private schools and inequality’, *Oxford Open Economics* **3**, 842–849.

Gregg, P., Washbrook, E., Propper, C. & Burgess, S. (2005), ‘The effects of a mother’s return to work decision on child development in the uk’, *The Economic Journal* **115**(501), 48–80.

Gupta, N. D. & Simonsen, M. (2010), ‘Non-cognitive child outcomes and universal high quality child care’, *Journal of Public Economics* **94**(1-2), 30–43.

Gupta, N. D. & Simonsen, M. (2016), ‘Academic performance and type of early childhood care’, *Economics of Education Review* **53**, 217–229.

Hansen, B. (2011), ‘School year length and student performance: Quasi-experimental evidence’. Social Science Research Networking Paper. Crossref.

Hansen, K. & Hawkes, D. (2009), ‘Early childcare and child development’, *Journal of Social Policy* **38**(2), 211–239.

Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (2004), ‘Disruption versus tiebout improvement: The costs and benefits of switching schools’, *Journal of Public Economics* **88**(9-10), 1721–1746.

Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (2009), ‘New evidence about brown v. board of education: The complex effects of school racial composition on achievement’, *Journal of Labor Economics* **27**(3), 349–383.

Havnes, T. & Mogstad, M. (2015), ‘Is universal child care leveling the playing field?’, *Journal of Public Economics* **127**, 100–114.

Heckman, J. J. (2011), ‘The economics of inequality: The value of early childhood education.’, *American Educator* **35**(1), 31.

Heckman, J. J. & Kautz, T. (2014), ‘The myth of achievement tests: The ged and the role of character in american life.’. Chicago: University of Chicago Press.

Heckman, J. J. & Rubinstein, Y. (2001), ‘The importance of noncognitive skills: Lessons from the ged testing program’, *American Economic Review* **91**(2), 145–149.

Hermansen, A. S. & Birkelund, G. E. (2015), ‘The impact of immigrant classmates on educational outcomes’, *Social Forces* **94**(2), 615–646.

Herrmann, M. A. & Rockoff, J. E. (2012), 'Worker absence and productivity: Evidence from teaching', *Journal of Labor Economics* **30**(4), 749–782.

Hewitt, R. (2020), 'Mind the gap: gender differences in higher education'. Accessed March 2024. `https://www.hepi.ac.uk/2020/03/07/mind-the-gap-gender-differences-in-higher-education/`,.

Hobbs, G. & Vignoles, A. (2010), 'Is children's free school meal 'eligibility'a good proxy for family income?', *British Educational Research Journal* **36**(4), 673–690.

Home Office (2020), 'Historical background information on nationality'. Accessed March 2024. `https://www.gov.uk/government/publications/historical-background-information-on-nationality/historical-background-information-on-nationality-accessible`.

Hoxby, C. M. (2000), 'Peer effects in the classroom: Learning from gender and race variation'. NBER working paper 7867. National Bureau of Economic Research Cambridge, Massachusetts, USA.

Huskinson, T., Pye, J., Medien, K., Dobie, S., Ferguson, C., Gardner, C., Gilby, N., Littlewood, M. & D'Souza, J. (2013), 'Childcare and early years survey of parents 2011'. Accessed: July 2024. `https://assets.publishing.service.gov.uk/media/5a7cbfc0e5274a38e575683a/SFR08-2013Text97-03_Updated_Jun13.pdf`.

Ilie, S., Sutherland, A. & Vignoles, A. (2017), 'Revisiting free school meal eligibility as a proxy for pupil socio-economic deprivation', *British Educational Research Journal* **43**(2), 253–274.

Ingram, J., Stiff, J., Cadwallader, S., Lee, G. & Kayton, H. (2023), 'Pisa 2022: National report for england. research report.', *UK Department for Education* .

Jackson, C. K. (2021), 'Can introducing single-sex education into low-performing schools improve academics, arrests, and teen motherhood?', *Journal of Human Resources* **56**(1), 1–39.

Jensen, P. & Rasmussen, A. W. (2011), 'The effect of immigrant concentration in schools on native and immigrant children's reading and math skills', *Economics of Education Review* **30**(6), 1503–1515.

Jessen, J., Spiess, C. K., Waights, S. & Wrohlich, K. (2021), *Sharing the caring? The gender division of care work during the Covid-19 pandemic in Germany*, IZA Discussion Paper.

Kamerman, S. B. & Waldfogel, J. (2005), 'Market and non-market institutions in early childhood education and care'. The limits of market organization (pp. 185–212). New York, NY: Russell Sage Foundation.

Keane, M. P. & Neal, T. (2021), 'A practical guide to weak instruments', *Annual Review of Economics* **16**.

Kleinjans, K. J. & Soest, A. V. (2014), 'Rounding, focal point answers and nonresponse to subjective probability questions', *Journal of Applied Econometrics* **29**(4), 567–585.

Kooreman, P. (2007), 'Time, money, peers, and parents; some data and theories on teenage behavior', *Journal of Population Economics* **20**, 9–33.

Kottelenberg, M. J. & Lehrer, S. F. (2014), 'Do the perils of universal childcare depend on the child's age?', *CESifo Economic Studies* **60**(2), 338–365.

Krueger, A. B. (2003), 'Economic considerations and class size', *The Economic Journal* **113**(485), 34–63.

Kuehnle, D. & Oberfichtner, M. (2020), 'Does starting universal childcare earlier influence children's skill development?', *Demography* **57**(1), 61–98.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E. & Liu, J. (2020), 'Projecting the potential impact of covid-19 school closures on academic achievement', *Educational Researcher* **49**(8), 549–565.

Lavy, V., Paserman, M. D. & Schlosser, A. (2007), 'Inside the black of box of ability peer effects: Evidence from variation in high and low achievers in the classroom', *NBER Working Paper W* **14415**.

Lee, D. S., McCrary, J., Moreira, M. J. & Porter, J. (2022), 'Valid t-ratio inference for iv', *American Economic Review* **112**(10), 3260–3290.

Legewie, J. & DiPrete, T. A. (2012), 'School context and the gender gap in educational achievement', *American Sociological Review* **77**(3), 463–485.

Lépine, A. & Estevan, F. (2021), 'Do ability peer effects matter for academic and labor market outcomes?', *Labour Economics* **71**, 102022.

Leuven, E., Lindahl, M., Oosterbeek, H. & Webbink, D. (2010), 'Expanding schooling opportunities for 4-year-olds', *Economics of Education Review* **29**(3), 319–328.

Lindqvist, E. & Vestman, R. (2011), 'The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment', *American Economic Journal: Applied Economics* **3**(1), 101–128.

Liu, J., Lee, M. & Gershenson, S. (2021), 'The short-and long-run impacts of secondary school absences', *Journal of Public Economics* **199**, 104441.

Lochner, L. (2020), 'Education and crime'. *The Economics of Education*, Academic Press 109–117.

Loeb, S., Bridges, M., Bassok, D., Fuller, B. & Rumberger, R. W. (2007), 'How much is too much? the influence of preschool centers on children's social and cognitive development', *Economics of Education review* **26**(1), 52–66.

Maestri, V. (2017), 'Can ethnic diversity have a positive effect on school achievement?', *Education Economics* **25**(3), 290–303.

Magnuson, K. A., Meyers, M. K. & Waldfogel, J. (2007), 'Public funding and enrollment in formal child care in the 1990s', *Social Service Review* **81**(1), 47–83.

Magnuson, K. & Duncan, G. J. (2016), 'Can early childhood interventions decrease inequality of economic opportunity?', *RSF: The Russell Sage Foundation Journal of the Social Sciences* **2**(2), 123–141.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *The Review of Economic Studies* **60**(3), 531–542.

Marcotte, D. E. (2007), 'Schooling and test scores: A mother-natural experiment', *Economics of Education Review* **26**(5), 629–640.

Marcotte, D. E. & Hansen, B. (2010), 'Time for school? when the snow falls, test scores also drop', *Education Next* **10**(1), 52–60.

Marcotte, D. E. & Hemelt, S. W. (2008), 'Unscheduled school closings and student performance', *Education Finance and Policy* **3**(3), 316–338.

Mas, A. & Moretti, E. (2009), 'Peers at work', *American Economic Review* **99**(1), 112–145.

Melhuish, E. (2004), 'A literature review of the impact of early years provision on young children, with emphasis given to children from disadvantaged backgrounds'. London, United Kingdom: National Audit Office.

Melhuish, E. & Gardiner, J. (2020), *Study of early education and development (SEED): Impact study on early education use and child outcomes up to age five years*, London: Department for Education.

Mendolia, S., Paloyo, A. R. & Walker, I. (2018), 'Heterogeneous effects of high school peers on educational outcomes', *Oxford Economic Papers* **70**(3), 613–634.

Menon, H. (2023), 'Childcare in the uk: Inefficient or underfunded?'. Socail Market Foundation. Accessed September 2023. `https://www.smf.co.uk/commentary_podcasts/childcare-in-the-uk-inefficient-or-underfunded/`.

Mora, T. & Oreopoulos, P. (2011), 'Peer effects on high school aspirations: Evidence from a sample of close and not-so-close friends', *Economics of Education Review* **30**(4), 575–581.

Morales, C. (2022), 'Do refugee students affect the academic achievement of peers? evidence from a large urban school district', *Economics of Education Review* **89**, 102283.

Morando, G. & Platt, L. (2022), 'The impact of centre-based childcare on non-cognitive skills of young children', *Economica* **89**(356), 908–946.

Morris, P. A., Connors, M., Friedman-Krauss, A., McCoy, D. C., Weiland, C., Feller, A., Page, L., Bloom, H. & Yoshikawa, H. (2018), 'New findings on impact variation from the head start impact study: Informing the scale-up of early childhood programs', *AERA Open* **4**(2), 2332858418769287.

Morris, S., Melhuish, E. & Gardiner, J. (2021A), *Study of early education and development (SEED): Impact study on early education use and child outcomes up to age three*, London: Department for Education.

Morris, T. T., Davey Smith, G., van Den Berg, G. & Davies, N. M. (2021B), 'Consistency of noncognitive skills and their relation to educational outcomes in a uk cohort', *Translational Psychiatry* **11**(1), 563.

Mueller, G. & Plug, E. (2006), 'Estimating the effect of personality on male and female earnings', *Ilr Review* **60**(1), 3–22.

Muñoz-Bullón, F., Sanchez-Bueno, M. J. & Vos-Saz, A. (2017), 'The influence of sports participation on academic performance among students in higher education', *Sport Management Review* **20**(4), 365–378.

Neymotin, F. (2009), 'Immigration and its effect on the college-going outcomes of natives', *Economics of Education Review* **28**(5), 538–550.

NHS (2023), 'Nhs history'. National Health Service. Accessed March 2024. https://www.england.nhs.uk/nhsbirthday/about-the-nhs-birthday/nhs-history/.

OECD (2023), 'Public spending on childcare and early education.'. Organisation for Economic Co-operation and Development. Accessed September 2023. https://www.oecd.org/content/dam/oecd/en/data/datasets/family-database/pf3_1_public_spending_on_childcare_and_early_education.pdf.

Office of National Statisitcs (2023A), 'Long-term international migration, provisional: year ending june 2023'. Accessed March 2024. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/longterminternationalmigrationprovisional/yearendingjune2023.

Office of National Statisitcs (2023B), 'Ethnic group, england and wales: Census 2021'. Accessed March 2024. https:https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021.

Office of National Statistics (2024), 'Population of england and wales'. Accessed March 2024. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest/#:~:text=By%20ethnicity%20over%20time%20(19%20groups),-Location%3A%20England%20and&text=Census%20data%20for%20England%20and,%2C%20from%201.2%25%20to%200.9%25.

Ofqual (2021), 'Learning during the pandemic: quantifying lost learning.'. London: Ofqual. Accessed July 2022. https://www.gov.uk/government/publications/learning-during-the-pandemic.

Oster, E. (2019), 'Unobservable selection and coefficient stability: Theory and evidence', *Journal of Business & Economic Statistics* **37**(2), 187–204.

Paloyo, A. R. (2020), Peer effects in education: recent empirical evidence, *in* 'The Economics of Education', Academic Press., pp. 291–305.

Parliament. House of Commons (1942), 'Ministry of health. volume 381: Debated on tuesday 30 june 1942'. Accessed March 2024. https://hansard.parliament.uk/Commons/1942-06-30/debates/07c05f13-740e-4e07-8861-ddebb766080c/MinistryOfHealth?highlight=nurse%20shortages#contribution-c152d47b-5894-4d9e-9ada-cc8985e8228c.

Pelletier, J. & Fesseha, E. (2019), 'The impact of full-day kindergarten on learning outcomes and self-regulation among kindergarten children at risk for placement in special education', *Exceptionality Education International* **29**(3), 42–56.

Peter, F. H., Schober, P. S. & Spiess, K. C. (2016), 'Early birds in day care: The social gradient in starting day care and children's non-cognitive skills', *CESifo Economic Studies* **62**(4), 725–751.

Pischke, J.-S. (2007), 'The impact of length of the school year on student performance and earnings: Evidence from the german short school years', *The Economic Journal* **117**(523), 1216–1242.

Plewis, I. (2007), 'Non-response in a birth cohort study: the case of the millennium cohort study', *International Journal of Social Research Methodology* **10**(5), 325–334.

Plucker, J. A., Eaton, J. J., Rapp, K. E., Lim, W., Nowak, J., Hansen, J. A. & Bartleson, A. (2004), 'The effects of full day versus half day kindergarten: Review and analysis of national and indiana data.', *Indiana University: Center for Evaluation and Education Policy* .

Rege, M., Telle, K. & Votruba, M. (2012), 'Social interaction effects in disability pension participation: evidence from plant downsizing', *The Scandinavian Journal of Economics* **114**(4), 1208–1239.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A. & Goldberg, L. R. (2007), 'The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes', *Perspectives on Psychological science* **2**(4), 313–345.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for dartmouth roommates', *The Quarterly Journal of Economics* **116**(2), 681–704.

Sacerdote, B. (2011), Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* 'Handbook of the Economics of Education', Vol. 3, Amsterdam: Elsevier, pp. 249–277.

Sammons, P., Toth, K., Sylva, K., Melhuish, E., Iraj, I. & Taggart, B. (2015), 'Pre-school and early home learning effects on a-level outcomes'. Effective Pre-school, Primary & Secondary Education Project (EPPSE). London: Department for Education.

Schneeweis, N. (2015), 'Immigrant concentration in schools: Consequences for native and migrant students', *Labour Economics* **35**, 63–76.

Schnepf, S. V. (2007), 'Immigrants' educational disadvantage: an examination across ten countries and three surveys', *Journal of Population Economics* **20**, 527–545.

Shao, X., Anders, J. & Macmillan, L. (2023), 'Persistent absenteeism: Who is missing school since the pandemic?'. London: Centre for Education Policy and Equalising Opportunities (CEPEO). Accessed March 2024. https://blogs.ucl.ac.uk/cepeo/2023/06/01/persistent-absenteeism-who-is-missing-school-since-the-pandemic/.

Sims, D. P. (2008), 'Strategic responses to school accountability measures: It's all in the timing', *Economics of Education Review* **27**(1), 58–68.

Snow, S. & Jones, E. (2011), 'Immigration and the national health service: putting history to the forefront.'. Institute of Historical Research, Senate House, University of London.

Sojourner, A. (2013), 'Identification of peer effects with missing peer data: Evidence from project star', *The Economic Journal* **123**(569), 574–605.

Speckesser, S. & Hedges, S. (2017), 'Peer effects and social influence in post-16 educational choice'. London: National Institute of Economic and Social Research.

Spence, M. (1973), 'Job market signaling the quarterly journal of economics, 87 (3)', *MIT Press* **355**, 374.

Spiess, C. K., Kreyenfeld, M. & Wagner, G. G. (2003), 'A forgotten issue: Distributional effects of day care subsidies in germany', *European Early Childhood Education Research Journal* **11**(2), 159–175.

Staiger, D. O. & Stock, J. H. (1994), 'Instrumental variables regression with weak instruments'. *Econometrica*, **65** 557–586.

Summers, A. A. & Wolfe, B. L. (1977), 'Do schools make a difference?', *The American Economic Review* **67**(4), 639–652.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. & Taggart, B. (2004), 'The effective provision of pre-school education (eppe) project technical paper 12: The final report-effective pre-school education'. Institute of Education, University of London/Department for Education and Skills.

Szulkin, R. & Jonsson, J. O. (2007), 'Ethnic segregation and educational outcomes in swedish comprehensive schools'. SULCIS reports and working papers, ISSN 1654-1189.

Taylor, L. (2023), 'Prolonged pandemic impact: School absenteeism at record levels.'. Accessed September 2023. https://www.brusselstimes.com/366824/prolonged-pandemic-impact-school-absenteeism-at-record-levels.

Todd, P. E. & Wolpin, K. I. (2003), 'On the specification and estimation of the production function for cognitive achievement', *The Economic Journal* **113**(485), F3–F33.

Tumen, S. (2021), 'The effect of refugees on native adolescents' test scores: Quasi-experimental evidence from pisa', *Journal of Development Economics* **150**, 102633.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N. & Network, N. E. C. C. R. (2010), 'Do effects of early child care extend to age 15 years? results from the nichd study of early child care and youth development', *Child Development* **81**(3), 737–756.

Votruba-Drzal, E., Li-Grining, C. P. & Maldonado-Carreño, C. (2008), 'A developmental perspective on full-versus part-day kindergarten and children's academic trajectories through fifth grade', *Child Development* **79**(4), 957–978.

Waldfogel, J. (1999), 'Early childhood interventions and outcomes'. CASE paper 21, Centre for Analysis of Social Exclusion, London School of Economics.

Whitney, C. R. & Liu, J. (2017), 'What we're missing: A descriptive analysis of part-day absenteeism in secondary school', *Aera Open* **3**(2), 2332858417703660.

Wiseman, J., Davies, E., Duggal, S., Bowes, L., Moreton, R., Robinson, S., Nathwani, T., Birking, G., Thomas, E. & Roberts, J. (2017), *Understanding the changing gaps in higher education participation in different regions of England*, England: Department for Education.

Xu, S., Xu, Z., Li, F. & Sukumar, A. (2021), 'Redefining peer learning: Role of student entrepreneurs in teaching entrepreneurship in the uk higher education context', *Industry and Higher Education* **35**(4), 306–311.

Younger, K., Gascoine, L., Menzies, V. & Torgerson, C. (2019), 'A systematic review of evidence on the effectiveness of interventions and strategies for widening participation in higher education', *Journal of Further and Higher Education* **43**(6), 742–773.

Zimmerman, D. J. (2003), 'Peer effects in academic outcomes: Evidence from a natural experiment', *Review of Economics and Statistics* **85**(1), 9–23.

Zvoch, K. (2009), 'A longitudinal examination of the academic year and summer learning rates of full-and half-day kindergartners', *Journal of Education for Students Placed at Risk (JESPAR)* **14**(4), 311–333.