Essays in Education Economics

Dominique A. Shure

Department of Economics and Lady Margaret Hall,

University of Oxford

A thesis under the supervision of Professor Margaret Stevens and Dr. Brian Bell,

submitted in fulfilment of the degree of

Doctor of Philosophy in Economics*

Trinity Term 2015

*Approximate word count: 77,000 words.

Abstract

This thesis examines three different aspects of education policy to ascertain their effects on individual outcomes, both in the classroom and in the labour market. The goal is to provide new empirical evidence using robust identification strategies that can inform better policy.

The first chapter looks at the role of pre-primary education in Germany using the German Socio-Economic Panel data set (GSOEP) to determine if attending an early education programme for longer increases the probability of attending a higher-level secondary school at age fourteen. I employ family fixed effects estimation and quasi-experimental analysis to control for selection. The results of the family fixed effects estimation show a small and negative impact of attending early education for more years. In the quasi-experimental analysis, based upon a federal law change in 1996, I find no impact of more years of early education on later schooling outcomes.

In the second chapter of this thesis, I again use the GSOEP to examine the recent German reform to extend the length of the primary school day. I exploit the quasi-experimental roll-out of reform to assign treatment to women and look at whether increasing school hours increases the likelihood that mothers enter into employment or extend their hours if already working. I find that the policy has an effect at the extensive margin, drawing more women into the labour market, but that there is no significant impact of the policy at the intensive margin.

In the final chapter I turn my attention to how peers' non-cognitive traits impact an individual's learning outcomes. Using an educational panel from Flanders, Belgium, I use the linear-in-means model of peer effects as well as several non-linear models to see how peers' personalities in a classroom affect Dutch and math scores. The results show that having more conscientious peers on average positively impacts Dutch and math scores, but that a greater dispersion of conscientiousness hurts Dutch outcomes. I also find that having more extroverted peers on average hurts math performance.

Acknowledgements

This thesis is the culmination of years of data entry, research, writing, and revisions. I am especially grateful for the continuous academic and personal support of my supervisors, Professor Margaret Stevens and Dr. Brian Bell.

Professor C. Katharina Spiess, Dr. Jan Marcus, and Dr. Jan Goebel and the German Institute for Economic Research (DIW) in Berlin also made this thesis possible by hosting me as a guest researcher on several occasions and by providing feedback and knowledge of the data. These visits would not have been possible without a Santander Travel Grant.

Professor Jan Van Damme and Bieke De Fraine from the KU Leuven provided immense support and help with the LOSO data set, answering many a question on content and structure. Dr. Kaska Kubacka also helped with my navigation of the LOSO data, the psychology literature, and Paris during my time at the Organisation for Economic Cooperation and Development.

I am especially grateful to the Weidenfeld-Hoffmann Trust for my five very generous years of funding, which enabled me to undertake this D.Phil. and enriched my Oxford experience greatly. All of the Trust's staff, and especially Lady Elizabeth Roberts proved to be great advocates of mine during the D.Phil.

My personal support system in Oxford and beyond also helped make this D.Phil. a reality. All of my fellow M.Phil. and D.Phil. students and post-docs in the department made surviving exams, thesis writing, presentations, and conferences a joy. I especially want to thank Abi Adams, Claudia Herresthal, Nick Chesterley, Michael Koelle, Galen Sher, Lucie Moore, Simon Quinn, Mathilde Almlund, Pavel Dziewulski, and Federico Torracchi. Their feedback and suggestions contributed to the quality of my research. I am most grateful to all the Unrewleys, who kept me sane at home, especially Dr. Elena Netsi, and all of my other non-economist friends, who still engaged with me on my thesis and in life, especially Mira Shastry, Kathryn Olivarius, and Joe Geylin. The support of my parents and sister, Torey, also kept me going through many a tough time. And last, but not least, I want to thank my partner, Thomas, my personal cheerleader and *Krafttier*.

Contents

1 Introduction

Education policy affects individuals and societies as a whole. Apart from the intrinsic benefit of education, governments invest in the education of their citizens because better educated citizens have been shown to exhibit more favourable behaviour, e.g. commit fewer crimes (Machin et al., 2011) and have better health outcomes (Campbell et al., 2014), and participate more productively in the labour market. Education provides skills and skills allow for specialisation and diversification of the economy. The linkage between education and the labour market is of interest to governments because human capital is one channel through which economic growth is achieved. Governments can use education policy to help shape the labour force they wish to have and thereby their economies.

At the individual level, the linkage between education and the labour market is also important. Individuals choose to invest in their own education, whether through taking on direct financial costs or opportunity costs, in order to obtain a payoff. Jacob Mincer popularised the idea that years of education could predict this payoff, future earnings, in what is now called the "Mincerian equation." These returns to education, both material (e.g. wages) and non-material (e.g. the prospect of a more educated spouse), entice individuals to invest. Investing in education can also be a means to social mobility. Despite the popularity of the Mincerian equation in empirical labour economics, we still do not fully understand all the mechanisms underlying it; this is especially true with regards to issues of endogeneity and heterogeneity. The questions of when to start this investment, who benefits most, and how what happens in the classroom affects the returns all remain open. In this thesis, I provide new evidence on some of these fundamental economics of education questions. Using longitudinal data sets from Germany and Belgium, I estimate the causal impact of specific education policies on an individual's labour market or academic outcome.

In the first chapter, I look at the question of how investing early in education affects later academic outcomes. The work of James J. Heckman has highlighted the importance of investing early, with claims that early education can increase future earnings, improve health outcomes, and make the achievement gap between low and high socio-economics status (SES) families smaller. The results of Heckman and co-authors' work has greatly shaped the policy debate in the U.S. and Europe, and placed a large focus on early childhood interventions. This dominant narrative, however, was created through a variety of high profile studies discussed in the first chapter, which do not reflect the reality of most early education programmes. While Heckman and co-authors focus on high cost, high quality, targeted interventions, most early education programmes in the U.S. and Europe are universal and heterogeneous in quality. I use data from Germany to explore the impact of attending a more "standard" early education programme on secondary schooling outcomes: are children who attend early education for longer more likely to end up at a better type of streamed secondary school? I also examine the claim that early education is especially beneficial for children from lower SES backgrounds by examining the impact of year of early education on children of immigrant families.

I use a family fixed effects approach to control for unobserved heterogeneity between families as well as a difference-in-difference estimation strategy exploiting a policy change that varied across states and time. I find no evidence that attending early education for longer increases the probability of attending a better secondary school. There is no statistically significant effect for children of immigrant families. In some instances, attending for longer actually reduces the likelihood of attending a better secondary school. This analysis does not take other outcome measures into account, due to data limitations, so these results do not rule out the potential benefit of early education in other areas of life. The lack of longer-term academic results shows, however, that policymakers need to be thinking not just about how to increase access to early education, but also about how to improve quality. The analysis in this chapter contributes pragmatic evidence that is more applicable to governments and parents in countries where universal early education is the norm.

In the second chapter of this thesis, I examine the relationship between childcare and maternal labour supply. Governments invest in education for the aforementioned reasons, including the fact that they want to create a better workforce; in the U.S. and Europe, men and women receive equal public investment in their primary and secondary education. From an economic perspective, the fact that women tend to work less over the lifecycle, primarily due to fertility decisions, means that a government's investment in men and women does not have the same average payoff in the formal labour market. If governments want to equalise the rates of return on their investment, then they need to enable women to enter or re-enter the labour market after having children. While most of the literature on this topic has focused on early childcare options, the length of the primary school day also greatly impacts a mother's ability to work. Because the length of the primary school day in most European countries does not match the length of the workday, parents, primarily mothers, often struggle to combine work and childrearing. Additional childcare outside of the public school setting may be limited and expensive. As will be discussed in the second chapter, this leads many women to work part time if they even work at all after having children.

The public discussion surrounding the issue of childcare and female labour supply has been increasing; David Cameron pledged to increase free childcare to 30 hours per week in the 2015 UK parliamentary elections, highlighting the rise of this issue in political discourse in Europe (Dominiczak, 2015). The length of the primary school day issue has been present in German consciousness due to one of the largest reforms to the German educational landscape in recent years: the *Ganztagsschulreform* or "full school day reform." This reform, partially aimed at increasing female labour supply, has affected thousands of pupils and their parents and received a considerable amount of media attention in Germany.

I use the reform to assess the impact of this implicit childcare subsidy on mothers' labour market participation. This type of reform is unusual and my analysis is some of the first to explore its impact at the extensive and intensive margins of the labour market. I use the quasi-experimental nature of the reform and geographical information software to assign treatment status to women and thereby estimate the causal impact of extending the school day. I am able to reconcile the results I obtain with a standard, static model of labour supply, with income and substitution effects. Understanding how these two effects interact to create labour supply outcomes is crucial in the policymaking context.

I find that women who have access to full day primary schools are more likely to enter into employment, but that women already working do not increase their hours. In fact, they are likely to decrease their working hours. Investing in childcare is an expensive policy for any government to undertake and policymakers want to be sure that their investment has the intended consequences, i.e. increasing labour supply. The evidence generated in this chapter can inform policymakers when thinking about how to best increase maternal labour supply, which is especially important as this issue gains political traction in other countries.

As the overview of these first two chapters shows, Germany proves an interesting country to study from an education and labour market perspective due to its complex education system and recent reforms. These two chapters are linked in that they explore the effects of childcare on children and their mothers. The nature of the German streamed education system and its importance in determining later life outcomes means that what type of school a child attends has long-term consequences. Children in most German states are sorted into secondary schools at the end of fourth grade, which compared to most other Organisation for Economic Cooperation and Development (OECD) countries is very early (OECD, 2004). This means that a pupil's education career up until university is practically locked in from the age of ten; this makes the type of secondary school a pupil attends all the more important and any measures to improve outcomes interesting.

At the same time, Germany has low female labour force participation, especially for mothers. In 2012, 68 percent of working age women were employed. Much of this employment, however, is part time. This is especially true when it comes to mothers as they face a short school day and limited and expensive formal childcare options. In 2002, only 16.8 percent of West German mothers and 51.7 percent of mothers in the East were engaged in full time employment (Wenzel, 2010). These statistics show why the German government would be interested in bringing more women into the workforce or helping them extend their hours if already working and why a policy to extend the school day could have an enormous impact.

How can we make sure that all children have the best opportunities available in a publicly funded education system? At the same time, how can we optimise school hours so that children learn as much as possible and work-life balance for parents becomes a reality? The analysis carried out in this thesis contributes to an existing evidence base using high quality German data combined with self-collected data never before used by other researchers. Although Germany has a unique education system, there are enough similarities with other OECD countries' systems to learn lessons from its education policy successes and failures.

In the final chapter of this thesis, I turn my attention to non-cognitive traits and peer effects. There has been growing research at the intersection of economics and psychology, especially in terms of non-cognitive traits (Borghans et al., 2008). Researchers are interested in expanding the traditional conception of the Mincerian equation to include non-cognitive traits because it is clear that more than years of education and work experience matter for labour market success. Economists can use aspects of psychological research to account for what would have previously been called "unobserved heterogeneity" between individuals. While we still might not be able to observe everything about a person, using psychological instruments such as personality tests and measures of motivation can greatly expand our existing knowledge. Again, Heckman has been leading the charge in this field, especially through his collaborations with psychologist Angela Duckworth. Their work has highlighted how important non-cognitive traits such as "grit" are in terms of economic outcomes (Borghans et al., 2008).

There is also a large body of research exploring how peers affect each other's outcomes; from smoking to teen pregnancy to academic and labour market outcomes, there is evidence that we are highly influenced by the people around us. In an education context, where group formation, i.e. classes, may be easily controlled, there is scope to positively impact outcomes. In Chapter 3, I use Belgian data on a cohort of secondary school pupils to examine the effect of who is in a classroom together, both in terms of academic and non-cognitive ability. Understanding how classmates affect each other can help schools optimise class allocation in order to maximise learning outcomes. Schools have traditionally tried to optimise allocation through methods of "streaming," whereby pupils of similar ability are grouped together in one class or possibly even a separate school. These kinds of efforts, however, have traditionally ignored other dimensions of pupils that may matter more for academic outcomes, e.g. personality. The data set I use contains psychological measures of all pupils in addition to their SES and academic performance. I am able to calculate class level measures of peer non-cognitive and cognitive measures and use these to see how an individual's learning outcomes are affected. I find that math and Dutch scores are affected differently by peer level non-cognitive traits. In the case of math scores, peer level extraversion and conscientiousness matter the most, while for Dutch, peer level conscientiousness matters. I also find that greater variance in a classroom's conscientiousness negatively impacts Dutch outcomes. Interestingly, I find that average peer level academic performance does not impact individual level performance, but I do find evidence of non-linear effects in the case of Dutch.

From an education policy perspective, the finding that personality influences outcomes proves challenging if we believe that personality is something static. The fact, however, that the composition of personalities within a group can also impact outcomes means that policymakers can sort pupils into groups in order to achieve a more optimal outcome. Grades no longer have to be the only parameter by which streaming occurs. This is a relatively low cost policy since it does not require any investment beyond collecting the necessary non-cognitive measures. Understanding the mechanisms at work in such a context is crucial to creating more effective learning situations and in this chapter, I aim to provide new evidence on this front.

This type of analysis, and much of the economics of education research in general, is plagued by issues of endogeneity in obtaining identification. Are more able children more likely to attend early education programmes and therefore also more likely to reach a better secondary school? Are there unobserved characteristics of families and children that fundamentally affect education and labour market outcomes? These are challenges I face throughout the thesis and in each chapter, I develop a creative and rigorous identification strategy based on my knowledge of education policy and the countries in question to address them.

The goal of this thesis is to use econometric modeling to shed light on some of the fundamental questions in the field of education economics and generate new evidence that may help create pragmatic policies. All three chapters aim to contribute to the existing evidence base on the success of policies and thereby deepen our understanding of the complex mechanisms underlying the relationship between education and the labour market.

2 The Returns to Early Education in Germany

2.1 Introduction

Does the early bird catch the worm when it comes to education? Does early education have a measurable impact on future outcomes? Early education programmes are lauded for the developmental advantages they foster amongst attendees; whether through formally acquiring a second language or simply learning how to count to ten, early education programmes offer varying degrees of pedagogical support. They also provide one of the first opportunities for children to socialise in an organised setting and develop inter-personal skills. Heckman has advocated focusing on investments in early child education rather than later interventions because the ones in early childhood matter so much more for future skill formation (Cunha and Heckman, 2007). If acquiring these various types of skills before enrolment in a formal primary school, or in a context outside of the home, gives these children an advantage in skill formation over their peers who do not attend such a programme, then early education matters.

Even if the merits of early education are acknowledged, the task of measuring its returns proves difficult. The effect of family background on schooling decisions is large. The Organisation for Economic Cooperation and Development's (OECD) programme for International Student Assessment (PISA) tests showed that in many of the OECD countries there is a strong correlation between test outcomes and socio-economic status (OECD, 2004). Becker (1983) postulated that education is an investment and that more educated people are more likely to give it a higher valuation and therefore invest more. This means that the families who send their children to an early education programme might have inherent characteristics that also cause their children to succeed later on in secondary school, confounding the effect of the early education programme.

Using German data, I will answer the question of whether or not early education impacts secondary schooling outcomes and disentangle the impact of family background from the effect of the early education programme. There has been little research on the returns to early education in Germany and its findings have been relatively inconclusive. Drawing on two different types of methodologies from the field of labour economics, I will build off of Spiess et al. (2003), one of the few German studies asking similar questions and using the same data, by focusing more explicitly on the underlying selection into education process.

My results using two different methodologies show that years of attendance in an early education programme has a small negative or no impact on secondary schooling placement for most children once I account for unobserved family heterogeneity.

2.2 Literature Review and Theoretical Motivation

2.2.1 Existing Literature

The existing literature on the returns to early education may be divided into studies that focus on targeted interventions and studies that look at universal programmes. Much of the existing literature on targeted interventions focuses on the U.S. context. Currie (2001) summarises the findings of some of the most important American studies. Some of the most noteworthy are the studies on the Carolina Abecedarian Project (Campbell and Ramey, 1994), the Perry Preschool Project (Schweinhart et al., 1993), the Chicago Child-Parent Centre (Fuerst and Fuerst, 1993), and Head Start (Currie and Thomas, 1995; Garces et al., 2002). All of these papers try to discern the impact of having attended an early education programme on future outcomes.

Schweinhart et al. (1998) analyse the effect of attending preschool on wage earnings at age twenty-seven using a Mincerian type equation and find positive effects on earnings as a result of having attended preschool. The intervention they study, the Perry Preschool Project, is often referred to as the "Cadillac" of early childhood education due to its high cost, and has shaped much of the literature and debate on early education in the U.S. and elsewhere.

Campbell and Ramey (1994) evaluate the effects of an experimental intervention, the Carolina Abecedarian Project, on low-income families. They find positive effects of early childhood education on academic outcomes through age twelve. Fuerst and Fuerst (1993) look at an intervention targeting low-income, mostly black children in Chicago and find larger benefits of attending for four to six years on reading scores once they enter primary school. These effects are larger for girls than for boys.

Currie and Thomas (1995) evaluate Head Start, a much less expensive, federally funded U.S. early education programme, using the National Longitudinal Survey's Child-Mother data set. They include mother fixed effects in order to control for selection into early education and find evidence of heterogeneous treatment effects along ethnic categories. White and Hispanic children have longer-term benefits on test scores and are less likely to repeat a grade than their African-American counterparts. In fact, they find no effect on the academic outcomes of African-American children who participated in Head Start. They also find evidence that white children who attended Head Start have higher test scores later in life than their siblings who did not attend Head Start. Garces et al. (2002) examine longer term benefits of Head Start using the Panel Study of Income Dynamics and find that white participants were more likely to graduate from high school and earn more money in their early 20s. They also find that African-American participants are less likely to be charged with or convicted of a crime. These results are important because they show that the effects of early education might not actually "fade out" as quickly as previously thought.

Much of the literature on the returns to early education has been dominated by the aforementioned studies on programmes such as the Perry Preschool Project or Carolina Abecedarian Project, which provided very high cost and high quality early education to a small, at-risk population. From a policymaking perspective, these studies are less informative if we are thinking about delivering early education to a much larger, heterogeneous group or even the whole population. Apart from Head Start, many of the U.S.-based programmes cannot realistically be scaled up due to their prohibitive costs. The type of early education I focus on in this chapter is also quite different from the previously mentioned studies in that it is widely available, heterogeneous in design, and not necessarily targeting a specific demographic group. This is why I now turn my attention to previous work exploring universal early education programmes.

Blanden et al. (2015) look at universal early education in the United Kingdom and provide an overview of other studies looking at universal programmes. In the UK, they find small, positive effects of universal early education on learning outcomes at age five; however, by age eleven, these effects have faded out and are no longer statistically significant. Berlinski, Galiani, and Gertler (2009) find a positive impact of expanding pre-primary education in Argentina on primary school outcomes and Berlinski, Galiani, and Manacorda (2008) find a positive effect of an expansion of early childhood education places on completed schooling in Uruguay.

Not all of the studies on universal early education find a positive effect on learning outcomes. In light of President Obama's "Preschool for All" initiative, Cascio and Schanzenbach (2013) discuss the effects of universal, state-funded early education programmes in Georgia and Oklahoma, which were implemented in the 1990s. They find that these programmes increased child preschool enrolment across socio-economic groups and that for low-income families, parents increased the amount of time they spent with their children. They do not, however, find any evidence that attending one of these early education programmes had any effect on test scores after grade 8. Similar to Blanden et al., these authors find evidence of a fade-out on academic outcomes. The fade-out explanation may also account for the findings in this chapter.

There have been several studies that explicitly look at returns to early education and the German education system. Dustmann et al. (2012) look at the same expansion of universal early childhood education based on a federal law change that came into effect on January 1, 1996 (see the Quasi-experimental Analysis section of this chapter for a further discussiong of this policy change). They focus on one state in Germany, Lower Saxony, and use administrative data. Dustmann et al. find that attending an early education programme improved school readiness and decreased motor and language problems for children of immigrant backgrounds, but did not affect native Germans. Their outcome measures are collected right before the children begin primary school, which is earlier than the outcome measures used in this chapter. These heterogeneous effects of early education by immigrant status is something I will explore in my analysis.

As part of a related literature, Puhani and Weber (2007) look at the effect of age of school entry on schooling outcomes at the end of primary school and in the middle of secondary school. They find that children who begin primary school at age 7 instead of 6 are 12 percentage points more likely to attend a *Gymnasium*. The authors acknowledge, however, that they are unable to identify the underlying mechanism driving this result, e.g. a relative age effect, a maturity effect, or a maturity-learning interaction, without more detailed data. Puhani and Weber also state that their findings are not evidence against early learning in general.

As previously mentioned, Spiess et al. look at whether or not attending an early education programme impacts the probability of attending one of the two upper forms of secondary school in Germany. I build on their work by asking the question if attending an early education programme in Germany increases the probability of attending a better type of German secondary school and if attending longer also affects secondary schooling placement. Here attending *Gymnasium* or *Realschule*, the two upper-tiered secondary schools in the streamed system, may be seen as the measure of success for the long-term benefits of attending an early education programme. Secondary school placement was also the outcome variable utilised by Spiess et al.[1]

Using the German Socio-Economic Panel data set (GSOEP), Spiess et al. attempt to demonstrate that attending an early education programme in Germany increases the probability of attending the two most elite forms of secondary school. They do not find statistically significant results to support their hypothesis, but they find some evidence that children of immigrant families who attend an early education programme are more likely to attend a better secondary school. Despite their efforts, however, they acknowledge that they cannot differentiate the treatment from the selection effect in their work.

As the discussion in this section has shown, the results of studies exploring the impact of early education on later academic outcomes is mixed. Some work in the German context has found returns early on and some evidence of heterogeneous treatment effects favouring children of immigrant backgrounds. I contribute to this literature by looking at the returns to early education in the middle of secondary school and using different methodological approaches to test the robustness of my results.

2.2.2 The Economics of Early Education

As in any returns to education framework, we conceptualise a return on the investment in early education later in life. The crucial difference in this case as compared to the standard returns to education framework (see Card (1999) for an overview) is that the individuals enrolled in education cannot make this participation or level decision themselves. Instead, parents must choose whether to send their children to early education, what type of early education, and how long they should

[1] I also look at the probability of attending just *Gymnasium*, but as these results do not differ from the combined schooling outcome measure, I do not present these results.

attend. This changes the nature of the investment and places a much larger importance on the constraints and decision-making process of the parents.

As with any investment, the investor will want to conduct a cost-benefit analysis before deciding whether or not to invest. In the case of investing in early childhood education there are costs to undertaking it. In terms of material costs, parents must be willing and able to pay the fees associated with an early education programme. These may vary depending on the context and be more or less affordable depending on how much the state subsidises it. There are also opportunity costs, both to the child and the parent. For both the child and the parent, there is the opportunity cost of foregone time spent together, which may also be an investment in later outcomes. The pedagogical quality of the parent's time spent with the child will depend on parental education level and socio-economic status.

As discussed in the literature review, there is evidence that many benefits may be accrued from investing in early education. The benefits to the child are the development of non-cognitive and cognitive skills, which impact later outcomes. These include socialisation with other children, learning self-control, learning to read and write, and other skills. The quality of the skills acquired will depend on the quality of the early education programme and the underlying ability of the child. There is some evidence that these benefits translate into a monetary reward in terms of higher future earnings; however, as previously discussed, much of this evidence comes from small, targeted, and costly interventions. Researchers are still not certain how likely these effects are to persist, as some evidence points to early returns on early education, which fade as time passes.

We would also expect to see different benefits based on characteristics of the child. From the peer effects literature (see Sacerdote (2011) for an overview), we know that where a child lies in the ability distribution affects learning outcomes. Children from high socio-economic backgrounds or of a high ability may actually receive lower quality input at an early education programme than they would at home. Conversely, children from disadvantaged backgrounds may benefit disproportionately from any sort of pedagogical input. These types of heterogeneous effects may occur across a variety of dimensions and should not be underestimated.

The heterogeneity between children in terms of ability might also impact the parents' decision to send their children to early education. Parents might decide to send their less able child to an early education programme for longer in order to give that child a boost before she enters primary school. Conversely, parents might decide to send their more gifted child for a longer period of time since this child is more likely to absorb more information and form stronger cognitive and non-cognitive skills. The direction of this relationship could clearly work in either direction.

For the parent there are also benefits to enrolling a child in early education. For mothers this benefit may be the possibility to re-enter the labour market after having children or extend hours if already working. Similarly, fathers might also benefit from the extended childcare option depending on the daily length of the programme and childcare dynamics within the family. Parents also benefit from having some of the education specific parental responsibilities shifted to childcare professionals, which may be more or less important depending on the education level of the parent.

This discussion illustrates the point that the constraints of the parents will play a fundamental role in the early education investment decision. Parents have both finite income and time and must choose an early education programme at the appropriate age for their child based on these constraints. This is quite a complex decision-making process that will also depend on place availability and the information parents have about various facilities. This decision-making process is also closely related to parents' labour market decisions. All of these factors make understanding the relationship between early education and later child outcomes all the more complex.

2.2.3 Empirical Strategy

In this chapter, ask the question of whether or not attending an early education programme for longer in Germany impacts secondary schooling outcomes. As discussed in the Existing Literature section, Spiess et al. use the German Socio-Economic Panel data set (GSOEP) to examine whether attending an early education programme in Germany increases the probability of attending the two most elite forms of secondary school. They use a sub-sample of the GSOEP to look at West Germany[2] only, which includes 316 pupils, and estimate a binary probit model. One of the biggest weaknesses with their paper is the failure to address issues of endogeneity. They do not use fixed effects or year dummies to control for unobserved heterogeneity, which limits the strength of their findings. They acknowledge that they cannot differentiate the treatment from the selection effect in their work.

I specifically address two shortcomings in the Spiess et al. methodology in this chapter: a limited sample size and an unresolved endogeneity problem. Because I am writing this chapter several years after Spiess et al., I am able to include more recent waves of the GSOEP in my analysis. I also expand my analysis to consider both East and West Germany, while Spiess et al. restrict their analysis to the West only. This increases the number of children I observe as compared to their paper. I also switch the focus on ever attending *Kindergarten* to the years of *Kindergarten* attendance since overall attendance at the extensive margin is fairly universal.

Dealing with an endogeneity problem is one of the largest challenges within the task of analysing the returns to early education. Here the endogeneity problem lies in unobserved heterogeneity. This unobserved heterogeneity occurs on both the child and family level. Unobserved heterogeneity between children arises because some children have a higher natural ability than others. In this

[2]For the purpose of this thesis, I use the term "West Germany" to refer to the western German states that used to form the Federal Republic of Germany from 1949-1990. I use this term even when referring to these states post-1990. The same logic applies to the eastern German states that made up the German Democratic Republic.

chapter, however, I am only able to control for potential differences between children based on gender and birth order. An ideal control variable would have been an IQ test score or some other aptitude measure, but my data set does not include information about this. For this reason, the rest of this chapter will attempt to tackle the issue of the unobserved heterogeneity between families.

All studies that deal with voluntary early education programmes must take into consideration that the students who attend such programmes may have other characteristics that contribute to their future success as well as explain why they attend such programmes in the first place. This means that I cannot simply estimate the Average Treatment Effect (ATE) of attending an early education programme since assignment to treatment is not random. A potential factor driving the selection into treatment may lie in unobserved heterogeneity between families and might include measures of socio-economic status such as parents' income and parents' education level; however, it might also include some non-measurable factors, such as a high valuation of education within the family. Spiess et al. attempt to control for this selection problem by including covariates for socio-economic status; they openly acknowledge, however, that this may not be enough to account for the unobserved heterogeneity. The goal of this chapter, then, is to build upon this work by finding solutions for this selection problem.

My methodology to account for the selection problem is drawn from previous work in education and labour economics. Currie and Thomas use a family fixed effects approach in their analysis of Head Start in the U.S. They look only at data on siblings where one sibling attended Head Start and the others did not. In my section on family fixed effects (FFE), I follow their methodology looking only at families where the siblings went to an early education programme for different numbers of years. This sample should allow me to use FFE to account for between family heterogeneity and selection into education based on these factors, whilst still examining the effect of attending an early education programme for different years. I compare my fixed effects estimates with my initial Ordinary Least Squares (OLS) estimates to establish consistency using an alternative method to calculate the fixed effects specification based on Mundlak (1978). I estimate all of the family fixed effects models using a linear probability model, which I discuss later on in the chapter.

After conducting the FFE analysis, I turn to a quasi-experimental methodology to account for the selection problem. I draw upon work by Card and Krueger (1994) and Campolieti and Riddell (2012), both of which employ a Difference-in-Difference-in-Differences (DDD), to assess the impact of a policy change. While neither of these papers explicitly tackle the question of an education policy change, both use geographical and time variation to assess the impact of a policy change on the people it induced into treatment.

I use an identification strategy based on an exogenous policy change, which opened up access to early education programme places, as an alternative way to deal with this selection problem. I then use difference analysis, both a Difference-in-Differences (DID) and a Difference-in-Difference-in-Differences, depending on the number of control groups that can be identified. My analysis in this section relies on comparing the means of a "treatment" and "control" group, which I identify based on how the exogenous policy shock affected people over time, regionally, or on the basis of some other dimension. I then compare the means of my outcome variable, for each group of children I identify, using the exogenous shock. This difference analysis may also be done using regression, enabling me to add additional control variables, which might also be driving the selection into education bias. Finally, I use instrumental variable regression to estimate the Local Average Treatment Effect (LATE) of expanded early education provision.

2.3 The German Education System: A Brief Overview

In this chapter, I use the German term *"Kindergarten,"* to refer to the German concept[3] of early education programmes (often called "nursery school" or "preschool" in other contexts) and not the American concept of "kindergarten."

Early education in Germany is voluntary and is designed for children until they enrol in mandatory primary school at age six. Many early education programmes in Germany are not state-run and charge a fee. In general, *Kindergarten* is regarded in Germany to be more than just childcare and also have an important educational aspect (Spiess et al., 2003). Within the data set I use, it is not possible to identify the type of early education programme a child attends or how much it costs; however, the heterogeneity of available programmes means that in most cases, families can find a programme to fit their needs and budget. This does present a challenge, however, in assessing the returns to early education since there is heterogeneity between programmes and quality might vary quite a lot. As in Spiess et al., I restrict my analysis to years of early education from age three onward since there is likely to be a larger pedagogical aspect to this early education as opposed to earlier years where the focus may be more care oriented.

There are four main types of secondary schools in the German education system: *Hauptschule, Realschule, Gymnasium*, and *Gesamtschule*. The first three are the schools that comprise the streamed system. The *Hauptschule* is the lowest level of secondary schooling and has an emphasis on vocational training; the *Realschule* is the middle or intermediate level of secondary schooling and focuses on preparing students for either a vocational school or a business-oriented school; the *Gymnasium* is the most elite of the four forms and provides the opportunity to earn a diploma called the *Abitur* that serves as the means of entry into the university system; and the *Gesamtschule* is the newest form of secondary school, provides all three forms of schooling in one building, and is most similar to an integrated secondary school (Robelen, 2005). Without the completion of the

[3]It should be noted that I am using *Kindergarten* to refer to all years of early education from age three to starting primary school although in German some of these years may actually be referred to as *Kita* or *Kindertagesstätte*.

Abitur, it is impossible to enter a German university. Because these two school forms afford pupils a range of opportunities from university education to apprenticeships in a variety of sectors, I treat attending either a *Gymnasium* or a *Realschule* as my measure of success. A student who attends a *Hauptschule* will have much more limited options in Germany, and less access to apprenticeship programmes, which correlates with lower future earnings (Prasad, 2004).

In this chapter, I focus on the three traditional forms of secondary schools, *Gymnasium, Realschule*, and *Hauptschule*, since they are comparable in terms of academic requirements and rigour and access to opportunities upon completion. The majority of secondary school age pupils in Germany attend one of these three forms of secondary schools, although there is a growing popularity in attending a *Gesamtschule.* Statistics from the German government indicate that in the 2010-2011 school year 11 percent of secondary school pupils in Germany attended a *Gesamtschule* (*Statistisches Bundesamt*, 2012). Due to the inability to rank the *Gesamtschule* in the same way as these three other school types, I exclude children who attended a *Gesamtschule* from my analysis.[4]

In Germany, a child begins secondary schooling at age ten in grade five.[5] The transition from a general primary school to a "streamed" secondary school takes place via the recommendation of the primary school teacher, which is binding in some German states, but can be contested by the parents in others. This recommendation can be based on a general assessment of the pupil or on actual performance with pre-determined cut-off points or some combination of the two. In every state there are actually different procedures that allow parents to overrule the teacher's decision; in some states parents can actually decide themselves (e.g. North-Rhine Westphalia) while in others parents must respect the decision of the teacher, but can ask to have their child try out a *Gymnasium* or a *Realschule* with the hope that the teachers there will accept her (e.g. Bavaria) (3sat, 2011). This variation across states and time could pose a problem in my estimation if it is systematic and ignored. I address this issue by including year dummies and a variable to capture the percent of pupils in a given state who attend a *Gymnasium* or a *Realschule.*

The dependent variable used in this chapter is a measure of secondary schooling placement at age fourteen instead of a more traditional option, such as wages, because of the way the data set is constructed. The adults interviewed answer questions about their children until the children reach the age of sixteen, at which point they are either interviewed separately or leave the survey.

[4]It is possible that there is some shared characteristic between pupils who attend a *Gesamtschule* and their families and that by excluding them, I am selecting my sample in a way that will affect my results. A series of t-tests on the means of key observable variables, e.g. parental education and income, between my sample and the excluded *Gesamtschule* pupils does not reveal any significant differences. It is still possible, however, that they differ on some unobservable traits, which would cause a problem for the representativeness of my sample. The number of pupils attending a *Gesamtschule* has increased over time, especially over the past decade; however, not all states have *Gesamtschule*, which means the proportion of pupils attending varies significantly across locations and time, only further complicating this issue. In all specifications I control for the percentage of pupils in a state attending a *Gymnasium* or a *Realschule* when the individual is 14, so that any fundamental differences in attendance at these two types of schools driven by *Gesamtschule* attendance over time will be taken into consideration.

[5]Except in the states of Berlin and Brandenburg, where pupils have the choice to begin secondary schooling at the end of 4th grade, but can also wait until the end of 6th grade.

This means that for any child who is age fourteen, I know what type of school they are enrolled in because their parents answer this question. Children in Germany are legally required to be enrolled in a secondary school at age fourteen, so this information should be available for all children in the data of this age.

My outcome variable is binary, taking the value one if the child is enrolled in either a *Gymnasium* or *Realschule* at age fourteen or zero if the child is enrolled in a *Hauptschule*. I use this variable because the labor market outcomes associated with attending either a *Gymnasium* or *Realschule* are favourable to attending a *Hauptschule*. People who attend a *Gymnasium* and complete *Abitur* may study at a university. People who complete the diploma at a *Realschule* are eligible to apply for a range of apprenticeships or technical courses at vocational tertiary institutions. As Prasad (2004) showed, there is a high degree of correlation between future earnings and having completed either *Gymnasium* or *Realschule*, which is why I choose to group these two school forms together. Spiess et al. used the same binary outcome variable grouping these two types of secondary schools together.

One thing to bear in mind when using this outcome variable is that it measures a relative outcome. Primary school teachers do not use quotas when they decide which type of secondary school a pupil will attend and parents have varying degrees of control over their child's secondary school placement. Theoretically, all children in Germany could attend either a *Gymnasium* or a *Realschule*, however, since resources are constrained, in practice it would be difficult. There will be some variation from year to year in how many children attend a *Gymnasium* or a *Realschule* depending on the ability distribution of the cohort, but it is unlikely to fluctuate in an extreme way. This means that even if going to an early education programme increases schooling outcomes, sending every child to an early education programme is unlikely to change the distribution of students who attend secondary schools due to space constraints. These constraints affect the effectiveness of the treatment and potentially create issues of cross contamination since many pupils who attend an early education programme might be forced to attend *Hauptschule* even if there educational outcomes have improved as a result of the treatment. This serves as an example of the type of interference that violates the stable unit treatment value assumption (SUTVA) and is something to keep in mind when analysing the results in this chapter.

2.4 Data and Descriptive Statistics

2.4.1 Data

In this chapter, I use the German Socio-Economic Panel Data Set (GSOEP, also known as SOEP) to construct a pooled cross-sectional data set specifically for my analysis. Every year since 1984, the German Institute for Economic Research (DIW) interviews members of approximately 11,000 households in Germany about a variety of issues. In this chapter, I use version 26 of the GSOEP, which spans the years 1984-2009 (SOEP v26). The issues covered by the GSOEP range from questions about the respondent's earning and employment to questions about their children and their children's education.

In the data, questions about the education of children are not asked retrospectively, meaning parents provide information about the current educational enrolment status of their children. For the purposes of this chapter, this means that for children who entered the survey once they were already enrolled in primary school or higher, I have no way of knowing whether or not they attended an early education programme. While this might be viewed as a potential downside due to the way it limits my sample size, Spiess et al. point out that this avoids the problem of "recall error." Because of this, I restrict my sample to include only children who entered the survey before age three. This means that all of the families represented in this data set did not attrite over at least a period of 11 years, which may mean there is something special about them as compared to the rest of the GSOEP and limit their representativeness. While these inclusion restrictions might create a sampling bias, it is a necessary to conduct this analysis.

The data set I create from the GSOEP has a total of 3,000 unique children. For the family fixed effects specification, I restrict this data set to only families with two children, giving me a sample size of 1,378 children. In some of the regressions presented later on, my sample size reduces below these numbers due to missing values in some of my covariates.

I construct a cross-sectional data set from the GSOEP instead of using the full panel of data because I am interested in the effect of a variable that does not change over time, years of early education attendance, on an outcome at a specific point in time, when the child is age fourteen; accordingly, I only observe the children at two points in their lives. This is the model of interest:

$$E_{ij} = x_i'\beta_1 + z_j'\beta_2 + \beta_3 yearskind_i + u_{ij} \qquad (1)$$

where the subscript "i" denotes the child, while "j" denotes the family. E_{ij} is my binary dependent variable taking the value one if the child is enrolled in a *Gymnasium* or a *Realschule* at age fourteen and zero if she is enrolled in a *Hauptschule* and $yearskind_i$ is my independent variable of interest that does not change over time.[6] In this data set, a child may attend *Kindergarten* for up to four years, so the variable $yearskind_i$ ranges from 0 to 4.

I extract data on all the children in my sample at age three, a reasonable age at which they would start an early education programme, but would not yet have started primary school, and again at age fourteen from the panel data set. I also extract some time-invariant information such as the maximum level of education in the child's household and the years of early education she

[6] I include a more detailed discussion of this model later on in the chapter.

attended. All of these children attend an early education programme between 1984-2001, which means that not only do they attend heterogeneous early education programmes, but they do so over an extended period of time. This may further contribute to the heterogeneity of early education experiences and is something I address in all of my models by including year dummies for when each child in the sample was fourteen years old.

I use the GSOEP for two main reasons. Firstly, it was until very recently one of the only representative panel data sets available on Germany. This meant it had a large sample size to offer and included some education relevant questions even though it is not specifically an education panel. It also includes information on all members in a household, including all siblings, which allows me to use family fixed effects. It was also the data set used by Spiess et al., one of the few papers to explore the relationship between early education and later schooling outcomes in the German context. Secondly, it is one of the few German data sets that allows me to identify the state in which people live. This proves crucial for the quasi-experimental analysis, in which I exploit state of residence to determine assignment to treatment. The newer National Education Panel (NEPS) is a much better data set on education measures and child specific factors, but does not identify state of residence or survey a child's siblings. This would not allow me to conduct the analysis undertaken in this chapter.

2.4.2 Descriptive Statistics

In Table 1, I provide a description of all of the variables used in my analysis, and in Table 2, I provide summary statistics for these variables. In Tables 4, 5, and 6, I look at several key socio-economic status variables for different demographic groups. I define the groups based on whether or not the child went to *Kindergarten*, is from an immigrant or non-immigrant family, or is East or West German because I want to see how my sample differs along these three dimensions.

In the full sample I use, 94 percent of the 3,000 children attended an early education programme and the average length of attendance was 2.16 years. This is similar to the West German average reported in Spiess et al. They report that the percentage of children attending an early education programme in Germany is 29.8 at age 3, 76.9 at age 4, and 94.8 at age 5 (Spiess et al., 2003). In Figure 1 I present the same information for my sample across all years. These numbers are slightly lower than those reported by Spiess et al., but I have included more children over a longer period of time where enrolment may vary more. This graph does not change much when I look at specific years. I present the distribution of years of attendance for the full sample in Figure 2.

Table 3 indicates that children from non-immigrant families are more likely to attend an early education programme than their counterparts from immigrant families and that East German children are also more likely to attend than children from the West. The difference between East

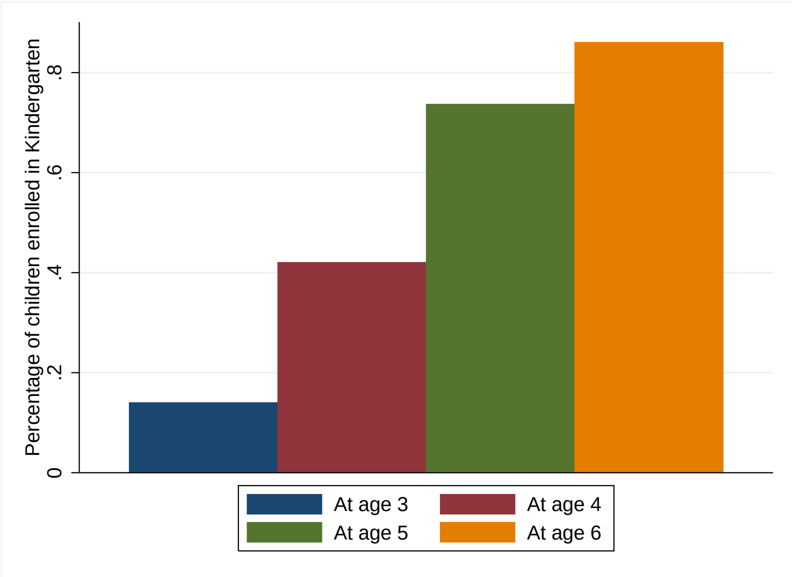


Figure 1: *Kindergarten* Attendance By Age

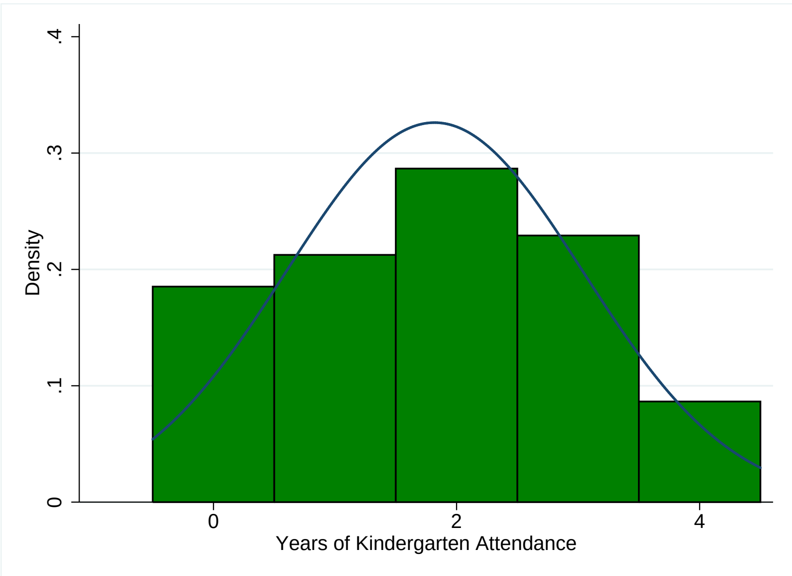


Figure 2: Distribution of *Kindergarten* Attendance

and West would be expected based the historical tradition of early education and childcare provision in former East Germany. The fact that children from immigrant families are less likely to attend also indicates that children from lower socio-economic status families are less likely to attend since the majority of immigrants in Germany, and in this data set, came to Germany under the guest worker programmes of the 1960s and generally fall at the lower end of the income distribution. Table 5 indicates that this is true for the immigrant families in my sample. They might be less likely to send their children to an early education programme because they are too financially constrained to afford the fees charged by many programmes or because they value education less than families of a higher socio-economic status.

In Table 4, I present the proportion of children from each of the demographic groups that attends a *Gymnasium* or a *Realschule* at age fourteen. Here, 63 percent of children who attended an early education programme go on to attend one of these better secondary schools versus only 48 percent of the children who do not attend such a programme. Breaking the children who attend an early education programme into categories based on the number of years of attendance shows that as the years increase, so too does the percentage of these children attending a *Gymnasium* or a *Realschule.* These descriptive statistics seem to support the idea that going to an early education programme increases the probability of attending a better secondary school and that length of attendance also has a positive impact.

For the other demographic groups, children from non-immigrant families are more likely to attend a *Gymnasium* or a *Realschule*, which could be explained by a linguistic advantage or higher socio-economic status. Table 4 also indicates that East German children are more likely to attend one of these two school types than West German children. This difference between East and West could be driven by an expanding number of places available in East Germany at these two types of schools over the period I am considering in my analysis, which would increase the probability of attending one of them. In order to make sure that the SUTVA is not violated as a result of this expansion, I include a control variable for the percentage of pupils in an individual's state who attend either a *Gymnasium* or a *Realschule* when the individual is 14 years old. This is something I will also take into account in the quasi-experimental analysis.

In Figure 3, I present data collected from the state statistical offices on the percentage of pupils enrolled at these two types of schools over time. I only present data from 1992-2009, since none of the children in my sample were enrolled in secondary school before 1992. Figure 3 shows an expanding percentage of students enrolled at these schools in East Germany, with a large decline in the early 2000s. This could be explained by "brain drain" from former Communist East Germany to more affluent areas of West Germany with better career opportunities, which has been a problem for the five former East German states post-reunification. Hunt (2000) looks at emigrants from East Germany post-reunification and finds that they are on average more skilled, and highly educated. If

Table 1: Variable Descriptions

Variable Name	Description
gymreal14	Binary variable for whether or not child in *Gymnasium* or *Realschule* at age 14
everkind	Binary variable for whether or not child attended *Kindergarten*
yearskind	Years a child attended an early education programme from age 3 onwards
male	Binary variable for gender
oldest	Binary variable for being the first born
maxhhedu	Highest number of years of education of either parent
migback	Binary variable for whether or not the child is an immigrant or the child of an immigrant
fulltime_m3	A binary variable for whether or not the child's mother was in full time employment when the child was age 3
fulltime_m14	A binary variable for whether or not the child's mother was in full time employment when the child was age 14
fulltime_d3	A binary variable for whether or not the child's father was in full time employment when the child was age 3
fulltime_d14	A binary variable for whether or not the child's father was in full time employment when the child was age 14
east_3	A binary variable for whether or not the child resided in East Germany at age 3
east_14	A binary variable for whether or not the child resided in East Germany at age 14
metro_3	A binary variable for whether or not the child resided in a metropolitan area at age 3
metro_14	A binary variable for whether or not the child resided in a metropolitan area at age 14
loginc_3	The logarithm of the household's monthly income post government transfers, normalised for number of people in household when the child was age 3
loginc_14	The logarithm of the household's monthly income post government transfers, normalised for number of people in household when the child was age 14
hhsize_3	The number of people living in the child's home when the child was age 3
hhsize_14	The number of people living in the child's home when the child was age 14
ssratio_14	The percentage of students in either a *Gymnasium* or a *Realschule* in the child's state when the child was age 14
hinckindavg	The average of the logarithm of the household's monthly income post government transfers, normalised for number of people in household while the child was enrolled in an early education programme
post96	A binary variable for whether or not the child was *Kindergarten* age before or after 1996
post98	A binary variable for whether or not the child was *Kindergarten* age before or after 1998
wnnrw	A binary variable for whether or not the child lives in West Germany, but not the state of NRW

Table 2: Summary Statistics

Variable Name	**Number of Observations**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
gymreal14	3,000	0.627	0.484	0	1
everkind	3,000	0.941	0.236	0	1
yearskind	3,000	2.159	1.092	0	4
male	3,000	0.511	0.500	0	1
oldest	3,000	0.451	0.498	0	1
maxhhedu	2,456	12.216	2.729	7	18
migback	3,000	0.300	0.458	0	1
fulltime_m3	2,011	0.147	0.354	0	1
fulltime_m14	3,000	0.236	0.424	0	1
fulltime_d3	2,011	0.844	0.362	0	1
fulltime_d14	3,000	0.741	0.438	0	1
east_3	2,011	0.160	0.366	0	1
east_14	3,000	0.184	0.388	0	1
metro_3	2,011	0.293	0.454	0	1
metro_14	3,000	0.266	0.442	0	1
loginc_3	1,892	7.382	0.523	3.084	9.682
loginc_14	3,000	7.659	0.507	3.537	10.224
hhsize_3	2,011	4.147	1.261	2	13
hhsize_14	3,000	4.269	1.173	2	14
ssratio_14	3,000	0.366	0.061	0.205	0.576
hinckindavg	2,723	7.426	0.448	5.213	9.108
post96	3,000	0.190	0.393	0	1
post98	3,000	0.190	0.392	0	1
wnnrw	2,448	0.723	0.447	0	1

Table 3: *Kindergarten* Attendance by Demographic Groups

	Attended an Early Education Programme			
	Mean	N	Standard Deviation	t-statistic
Immigrant Background				
Has an immigrant background	0.898	901	0.303	
Does not have an immigrant background	0.960	2,099	0.197	
				6.611 (0.000)
East vs. West German				
East German	0.960	552	0.196	
West German	0.937	2,448	0.244	
				-2.114 (0.035)

NB: p-value presented in parentheses below t-statistic

this is a predictor of their children's schooling outcomes, then a mass exodus of skilled people from East Germany could be driving the approximate 10 percent decline in *Gymnasium* and *Realschule* pupils. Because of this change over time, I control for the percentage of pupils enrolled in these types of schools in my analysis.

In Table 5, I compare my sample in terms of household income and parental education. This shows that children who attend an early education programme are more likely to come from wealthier families and have more educated parents than those who do not. The parents of the children who did not attend Kindergarten have an average of 10.836 years of education, which is significantly lower than the average for children who attended, 12.408 years. Again, this supports the ideas of Becker (1983), specifically that higher socio-economic status families make a larger investment in education. I also look at the children who attend an early education programme based on the number of years of attendance. Table 5 shows a relatively small range of average log household incomes for any number of years of attendance, while average parents' education increases with the number of years of attendance. This would imply that families who value education, i.e. the parents who have a high education level, are more likely to invest in early education and send their children for longer. Table 5 also indicates that immigrant children tend to come from more disadvantaged backgrounds than their non-immigrant counterparts. Interestingly, the difference between East and West in my sample is almost non-existent in terms of average household income, but does arise in terms of education.

Table 4: Secondary Schooling Outcomes by Demographic Groups

	Attended a *Gymnasium* or *Realschule* at Age 14			
	Mean	N	Standard Deviation	t-statistic
***Kindergarten* Attendance**				
Went to *Kindergarten*	0.636	2,823	0.481	
Did not attend *Kindergarten*	0.480	177	0.501	
				-4.173 (0.000)
Years of *Kindergarten* Attendance				
1 year	0.604	560	0.490	
2 years	0.632	1,023	0.482	
3 years	0.638	814	0.481	
4 years	0.714	220	0.453	
				10.390 (0.000)
Immigrant Background				
Has an immigrant background	0.486	901	0.500	
Does not have an immigrant background	0.687	2,099	0.464	
				10.646 (0.000)
East vs. West German				
East German	0.77	552	0.421	
West German	0.595	2,448	0.491	
				-7.761 (0.000)

NB: p-value presented in parentheses below t-statistic

Table 5: Socio-Economic Summary Statistics by Demographic Groups

	Household Income at Age 3				Parents' Education			
	Mean	N	SD	t-stat	Mean	N	SD	t-stat
***Kindergarten* Attendance**								
Went to *Kindergarten*	7.395	1,823	0.522		12.408	2,313	2.729	
Did not attend *Kindergarten*	7.035	69	0.411		10.836	143	2.268	
				-5.512 (0.000)				-6.746 (0.000)
Years of *Kindergarten* Attendance								
1 year	7.253	213	0.487		11.989	464	2.758	
2 years	7.396	655	0.523		12.428	856	2.675	
3 years	7.448	652	0.494		12.487	654	2.670	
4 years	7.349	159	0.639		12.769	175	2.789	
				12.580 (0.000)				23.380 (0.000)
Immigrant Background								
Has an immigrant background	7.224	600	0.542		10.906	760	2.449	
Does not have an immigrant background	7.455	1,292	0.497		12.948	1,696	2.610	
				8.763 (0.000)				18.271 (0.000)
East vs. West German								
East German	7.329	207	0.440		13.323	451	2.443	
West German	7.389	1,685	0.532		12.090	2,005	2.739	
				1.149 (0.251)				-8.800 (0.000)

NB: p-value presented in parentheses below t-statistic

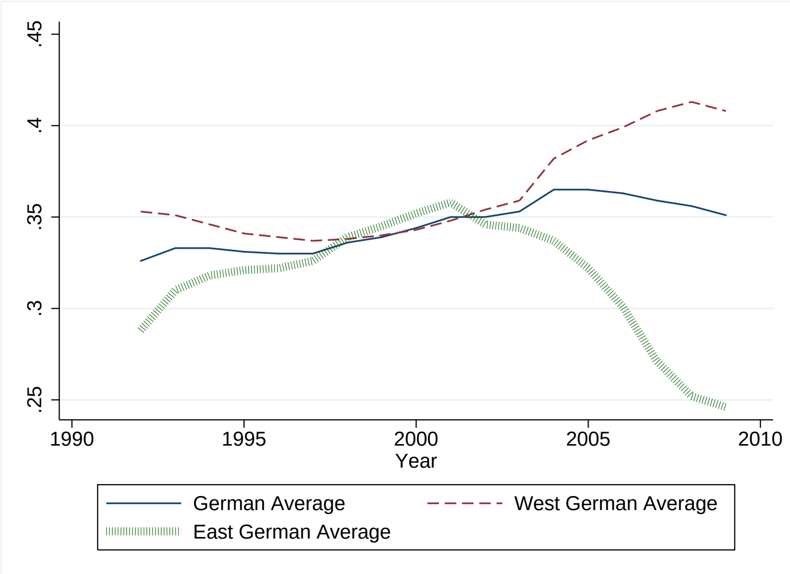


Figure 3: *Gymnasium* and *Realschule* Attendance 1992-2009: German Averages

Table 2 is also informative about the employment status of the parents of the children in my sample. While the majority of fathers of the children in my sample are in full time employment when the child is both age three and age fourteen, the percentage of mothers in full time employment at either point in time is much lower. I would expect the employment status of the mother to be highly correlated with *Kindergarten* attendance since early education programmes also serve a childcare purpose. Only 14.7 percent of mothers in this sample are in full time employment when their children are three years old, something that could point to later entry into an early education programme if these programmes are primarily used to provide childcare for working parents. The percentage of working mothers increases to 23.6 percent once the children turn fourteen, implying that mothers are re-entering the labour force as their children get older. This statistic ignores mothers who work part time and might also use an early education programme as a source of childcare.

I examine the mothers' propensity to work full time in Table 6 for the different demographic groups. When looking at whether or not the mother worked when the child was age 3 and attended an early education programme, it becomes apparent that children who do not attend an early education programme do not have a mother who is employed full time. Instead, it is likely that childcare is being provided in the home; whether or not early education is being provided in the home is a different issue. Table 6 also shows that children from immigrant families as well as those from East Germany are more likely to have mothers who work full time, both of which could be explained by family structures and gender roles in East Germany as well as socio-economic status.

Overall these descriptive statistics indicate that the families who invest in early education are more likely to be of a higher socio-economic status, both in terms of education and income, which implies a selection problem based on family characteristics. This is something I will test for in the fixed

Table 6: Mothers' Propensity to Work Full Time by Demographic Groups

	Mother in Full Time Employment at Age 3				**Mother in Full Time Employment at Age 14**			
	Mean	N	SD	t-stat	Mean	N	SD	t-stat
***Kindergarten* Attendance**								
Went to *Kindergarten*	0.152	1,940	0.359		0.240	2,823	0.427	
Did not attend *Kindergarten*	0.028	71	0.167		0.164	177	0.371	
				-2.887 (0.004)				-2.322 (0.020)
Immigrant Background								
Has an immigrant background	0.167	603	0.374		0.239	901	0.426	
Does not have an immigrant background	0.138	1,408	0.346		0.234	2,099	0.424	
				-2.322 (0.020)				-0.250 (0.803)
East versus West German								
East German	0.371	318	0.484		0.449	552	0.498	
West German	0.105	1,693	0.307		0.188	2,448	0.390	
				-12.763 (0.000)				-13.477 (0.000)

NB: p-value presented in parentheses below t-statistic

effects portion of this chapter.

I also want to know how the children who attend an early education programme differ from those who do not on a regional basis. In Table 7, I present information on what percentage of children who attend an early education programme live in an urban or metropolitan area at age three versus those who do not attend and how many of them are from East Germany. I define the dummy variable "metro" as taking the value one if the child lives in a community with a population of more than 100,000 people. A high percentage of children who come from an urban area do not attend an early education programme, which might be driven by socio-economic status or access to alternative forms of childcare. I also look at how children who attend an early education programme differ along these two dimensions depending on the number of years they attend. This table reveals that a majority of children who attend an early education programme for three or more years come from East Germany. This implies that attitudes and behaviour regarding early

Table 7: Regional Variation and *Kindergarten* Attendance

	Lived in metropolitan area				**Lived in East**			
	Mean	N	SD	t-stat	Mean	N	SD	t-stat
***Kindergarten* Attendance**								
Went to *Kindergarten*	0.288	1,940	0.453		0.160	1,940	0.367	
Did not attend *Kindergarten*	0.423	71	0.497		0.169	91	0.377	
				2.457 (0.014)				0.208 (0.835)
Years of *Kindergarten* Attendance								
1 year	0.335	215	0.473		0.047	215	0.211	
2 years	0.276	663	0.447		0.033	663	0.179	
3 years	0.282	674	0.450		0.089	674	0285	
4 years	0.274	212	0.447		0.462	212	0.500	
				2.140 (0.073)				215.980 (0.000)

NB: p-value presented in parentheses below t-statistic

education in East versus West Germany differ.

It is also interesting to see how early education programme attendance changes over time within my data set, especially for my later analysis using a policy change that expanded access starting in 1996. I use my data to calculate the percentage of children under age six attending an early education programme in each year of the survey and present this information in Figure 4. I only include data up until 2001, since I do not have secondary schooling outcomes data at age fourteen for children who attended an early education programme past this date. This figure shows a relatively constant trend until 1990, when Germany reunified, with an increase as a result of the traditionally higher attendance rates in East Germany. This higher level also remains constant until 1996, when a large increase occurs. This is likely due to a federal law change in 1996, which guaranteed all children from age three until they start primary school a place at an early education programme. I will estimate the effect of this policy change on the children in my sample in the quasi-experimental section of this chapter.

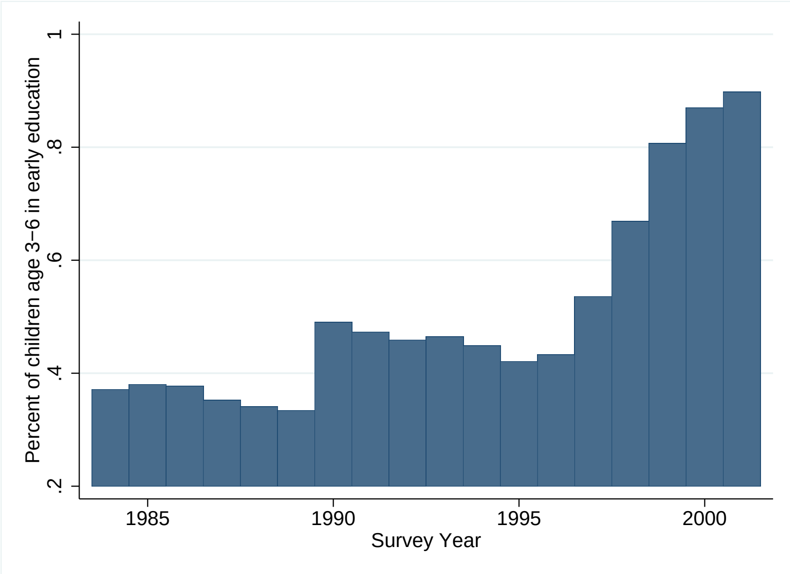


Figure 4: Early Education Attendance for Children Age 3-6 in Sample, 1984-2001

2.5 Initial OLS Results

To complement my descriptive statistics, I estimate a standard ordinary least squares (OLS) regression on my full sample. I estimate the same linear probability model previously discussed in the Empirical Strategy section as Equation (1):

$$E_{ij} \quad = \quad x_i'\beta_1 + z_j'\beta_2 + \beta_3 yearskind_i + u_{ij}$$

Table 8 shows the OLS results from regressing the secondary schooling outcome variable on just the dummy variable for ever having attended *Kindergarten* in Column (1) and the years of *Kindergarten* variable in Column (2). Both of these specifications include year dummies, which capture any cohort effects of having attended secondary school in a given year. I also use heteroskedasticity robust standard errors to account for the heteroskedasticity introduced through the OLS estimation of a binary outcome model. Both of the coefficients on the variables for ever having attended *Kindergarten* and the years of *Kindergarten* attendance are positive and statistically significant, indicating that there is a positive correlation between attending an early education at all and attending for increasing years and secondary schooling placement. These results do not differ when I use a probit instead of OLS to estimate these models.

The results of the full linear probability model estimated by OLS on the entire sample are presented in Table 9. These regressions include covariates measured when the child is three years old and when she is fourteen years old. They indicate a small, positive, and insignificant association between ever attending an early education programme (Column (1)) and the years of *Kindergarten* attendance (Column (2)) and secondary schooling outcomes. These baseline estimates are in line with what might be expected: early education matters in a small, positive way. This corresponds to the earlier

Table 8: OLS Mean Estimates on Full Sample

VARIABLES	(1) gymreal14	(2) gymreal14
everkind	0.147***	
	(0.043)	
yearskind		0.047***
		(0.010)
Year dummies	Yes	Yes
Constant	0.464***	0.495***
	(0.056)	(0.043)
Observations	3,000	3,000
R-squared	0.029	0.034

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

descriptive statistics showing that a higher percentage of children who attend an early education programme end up at a *Gymnasium* or a *Realschule*.

I look for the possibility a non-linear relationship between years of early education and secondary schooling outcomes in Columns (3) and (4), but do not find any evidence to support this. All of the coefficients on the dummies for individual years of overall attendance are positive and statistically insignificant. When I turn my attention to dummies for attendance at a certain age, I find a small, negative effect on the dummy variable for attendance at age three, but this is not statistically different from zero. This points to the lack of a non-linear relationship across years of attendance meaning that different years of attendance do not have a different impact on the outcome measure.

I also look at heterogeneous associations between early education and secondary schooling outcomes in this OLS framework to see if I find support for any of the mechanisms discussed earlier. The results estimated in Columns (1) and (2) are average results across the entire sample and I want to see if these associations differ on the basis of certain characteristics of the child or its family. The results in Column (5) show that years of early education attendance has a statistically significant and positive effect for girls and a zero effect for boys. There is substantial evidence from developmental psychologists that children of different genders develop differently and this could explain why they react differently to early education. Blanden et al. (2015) point out that most studies in this literature find greater returns to early education for girls than for boys. The findings in this chapter may be reconciled with much of the literature.

Column (7) shows a very large and statistically significant impact of years of early education attendance for children with less educated parents. The impact decreases as years of education of the parents decreases. This points to the mechanism discussed earlier that early education may have a larger impact on children who come from disadvantaged backgrounds and that for children from educated families, there can actually be a negative impact of being in a programme with less input than they would have received at home. The interaction terms for oldest child in Column

(6) and immigrant family in Column (8) do not reveal any statistically significant heterogeneous effects of early education along these dimensions.

This model does not, however, account for the unobserved family characteristics that might be driving selection into either early education or better secondary education. The descriptive statistics have shown that families of a higher socio-economic status are more likely to send their children to an early education programme and that they are more likely to attend a *Gymnasium* or *Realschule* at age fourteen. Can we disentangle the effect of the early education programme from the selection based on socio-economic status or other family-specific factors? To answer this question, I now turn my attention to family fixed effects estimation on the siblings-only sub-sample of my data, and quasi-experimental analysis to estimate the Local Average Treatment Effect (LATE).

Table 9: Initial OLS Estimates on Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14
everkind	0.053							
	(0.066)							
yearskind		0.011			0.027*	0.020	0.170***	0.000
		(0.012)			(0.016)	(0.015)	(0.045)	(0.015)
1.yearskind			0.055					
			(0.061)					
2.yearskind			0.069					
			(0.057)					
3.yearskind			0.066					
			(0.057)					
4.yearskind			0.072					
			(0.064)					
kindage3				-0.011				
				(0.034)				
kindage4				0.008				
				(0.025)				
kindage5				0.003				
				(0.031)				
kindage6				0.052				
				(0.035)				
yearskind*male					-0.032*			
					(0.020)			
yearskind*oldest						-0.022		
						(0.021)		
yearskind*maxhhedu							-0.013***	
							(0.003)	
yearskind*migback								0.028
								(0.024)
male	-0.081***	-0.082***	-0.082***	-0.082***	-0.004	-0.081***	-0.083***	-0.080***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.052)	(0.021)	(0.021)	(0.021)
oldest	0.014	0.015	0.015	0.015	0.015	0.068	0.013	0.017
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.057)	(0.023)	(0.023)
maxhhedu	0.049***	0.049***	0.049***	0.049***	0.049***	0.049***	0.082***	0.049***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.009)	(0.005)
migback	-0.041	-0.041	-0.039	-0.039	-0.043	-0.041	-0.035	-0.104*
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.061)
fulltime_m3	-0.000	-0.003	-0.003	0.001	0.000	-0.002	0.003	-0.004
	(0.036)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
fulltime_m14	-0.068**	-0.068**	-0.067**	-0.068**	-0.067**	-0.067**	-0.064**	-0.067**
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
fulltime_d3	0.073*	0.074*	0.073*	0.072*	0.074*	0.074*	0.075*	0.075*
	(0.039)	(0.039)	(0.039)	(0.039)	(0.038)	(0.039)	(0.039)	(0.039)
fulltime_d14	0.102***	0.102***	0.102***	0.101***	0.102***	0.101***	0.104***	0.102***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
east_3	0.157	0.145	0.154	0.156	0.141	0.147	0.150	0.155
	(0.106)	(0.106)	(0.108)	(0.107)	(0.104)	(0.103)	(0.105)	(0.107)
east_14	0.033	0.036	0.034	0.033	0.041	0.034	0.041	0.033
	(0.103)	(0.101)	(0.102)	(0.101)	(0.100)	(0.099)	(0.100)	(0.102)
metro_3	-0.026	-0.025	-0.026	-0.027	-0.028	-0.024	-0.015	-0.025
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
metro_14	0.028	0.027	0.029	0.030	0.030	0.026	0.026	0.030
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.046)	(0.047)
loginc_3	0.078**	0.078**	0.079**	0.077**	0.077**	0.078**	0.076**	0.077**
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
loginc_14	0.061**	0.061*	0.060*	0.062**	0.061*	0.062**	0.060*	0.061*
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
hhsize_3	-0.004	-0.004	-0.003	-0.003	-0.004	-0.004	-0.002	-0.003
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
hhsize_14	-0.015	-0.015	-0.015	-0.015	-0.016	-0.015	-0.016	-0.015
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
ssratio_14	0.133	0.135	0.119	0.108	0.141	0.133	0.110	0.129
	(0.288)	(0.287)	(0.288)	(0.288)	(0.287)	(0.287)	(0.287)	(0.287)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-1.207***	-1.180***	-1.210***	-1.207***	-1.214***	-1.210***	-1.556***	-1.148***
	(0.309)	(0.308)	(0.309)	(0.306)	(0.309)	(0.312)	(0.325)	(0.308)
Observations	1,834	1,834	1,834	1,834	1,834	1,834	1,834	1,834
R-squared	0.189	0.189	0.190	0.190	0.190	0.190	0.194	0.190

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.6 Family Fixed Effects

I begin my analysis in this section of the chapter by defining the linear probability model I wish to estimate. Because most children in my sample attended an early education programme, I focus on the number of years they attended. This differs from Spiess et al., which looks at the extensive margin of attending versus not attending. I begin with a random effects, OLS estimation on the siblings-only sample and then turn to family fixed effects estimation. I compare the random effects estimates on this sample with the family fixed effects estimates using a Mundlak (1978) test. I conclude this section by examining the scope for non-linearities in the relationship between years of early education and secondary schooling outcomes and heterogeneous treatment effects.

2.6.1 The Linear Probability Model

I define the following linear probability model (LPM) for the binary secondary schooling outcome,

$$E_{ij} = x_i'\beta_1 + z_j'\beta_2 + \beta_3 yearskind_i + u_{ij} \tag{2}$$

$$u_{ij} = \eta_j + \varepsilon_i \tag{3}$$

where the subscript “i” denotes the individual child, while “j” denotes the family. The error term, u_{ij}, is composed of a family-specific time-invariant component, η_j, and an individual-specific time-invariant component, ε_i. This LPM also has a matrix of child specific regressors, x_i, as well as a matrix of observed family-specific regressors, z_j. E_{ij} is my binary dependent variable taking the value one if the child is enrolled in a *Gymnasium* or a *Realschule* at age fourteen and zero if she is enrolled in a *Hauptschule*. I use an LPM because of its ease of interpretation and prevalence in the literature (Quinn, 2010). Although the LPM has some well-known shortcomings, e.g. probabilities may fall outside of the range [0, 1], it also allows for straightforward interpretation of its results.

For the variables in the matrix z_j, which in my analysis includes the maximum education of either parent and whether or not the child comes from an immigrant family, children from the same family will have the same observed value. The x_i matrix includes variables measured at age three and age fourteen for the individual child: log household income, a dummy variable for East Germany, a dummy variable for living in a metropolitan area, full time employment status of the mother and father, the size of the household, and the percentage of pupils in the child's state enrolled in a *Gymnasium* or a *Realschule*. The x_i matrix also includes dummy variables for whether or not the child is a boy or the oldest, as well as the family's income during the time before the child enrolled in primary school. Even though some of these variables seem family-specific, because they are also specific to the time when the child was either three or fourteen, they vary across siblings since all of these siblings are different ages. A complete list with a description and summary statistics of these covariates is provided in Tables 1 and 2.

This identification strategy relies on the assumption that years of early education programme attendance is uncorrelated with unobserved individual effects, but is correlated with unobserved family effects. This means that families choose the number of years they send their children to an early education programme based on unobservable family-specific characteristics, such as how much the family values education.

$$E(u_{ij} \mid \quad x_i, z_j, yearskind_i) \neq \quad 0 \tag{4}$$

$$E(\varepsilon_i \mid \quad x_i, z_j, yearskind_i) = \quad 0 \tag{5}$$

$$E(\eta_j \mid \quad x_i, z_j, yearskind_i) \neq \quad 0 \tag{6}$$

The problem in estimation arises due to the endogeneity present in the model as a result of the unobserved family-specific time-invariant effects. Assuming unobserved individual effects are uncorrelated with the regressors is a strong assumption and will be discussed later on with the results.

2.6.2 Family Fixed Effects Estimation on the Siblings-Only Data

In this section of the chapter, I deal with non-randomised assignment to treatment by differencing out the time-invariant family characteristics, which I believe to be driving the selection into education process. Currie and Thomas face a similar challenge in their analysis of the U.S. early education programme Head Start. The Head Start programme is different than the early education programmes being assessed in this chapter, as Head Start is a standardised programme that specifically targets low-income families. Nevertheless, Currie and Thomas face a similar selection bias challenge since Head Start is also a voluntary programme. In order to deal with this selection, they use a sample consisting exclusively of children with siblings, where one sibling attended Head Start and the others did not. In their sample, the siblings who did not attend Head Start either did not attend an early education programme or attended a non-Head Start programme. Currie and Thomas then use family fixed effects regressions, controlling for child-specific variables, on their siblings-only sample.

In the data used in this chapter, there are not enough multi-sibling families where one child attended an early education programme and the others did not (or at least one did not). Because of this, I turn my attention to families where the years of attendance differ between the siblings within one particular family. Identification in this case relies on there being enough variation in years of early education attendance within families once I have controlled for household income and parental employment when the child is three.

Table 10: Years of Attendance by Siblings

		Younger Sibling Number of Years					
		0	**1**	**2**	**3**	**4**	**Total**
	0	2	17	6	6	1	32
	1	50	2	59	37	14	162
	2	27	26	4	91	31	179
Older Sibling Number of Years	**3**	11	12	68	9	24	124
	4	2	6	6	19	19	52
	Total	92	63	143	162	89	549

NB: All numbers in table represent pairs of siblings

I only look at families with two siblings, just as Currie and Thomas (1995), in order to eliminate any dynamics that might be present with birth order in a multi-sibling family. This gives me 1,098 children from 549 families.

In Table 10, I compare the sibling pairs in terms of years of early education attendance. As this table shows, there is enough variation in years of attendance between sibling pairs to use FFE to obtain robust estimates. It is not the case that most siblings attend for the same number of years or that either all younger or older siblings attend for a longer period of time.

As Currie and Thomas point out, household or family fixed effects control for time-invariant characteristics of the household such as the permanent income and education levels of the parents, as well as constant unobservable taste and preference measures. The additional child-specific variables they include are: age, gender, a dummy for being the first born, and household income during the time the child was enrolled in either Head Start or another programme. They are also interested in the difference between permanent family income and the family's income when a child is three years old, as this particular income level can vary between siblings and might also explain why siblings go to different early education programmes, or in my example, for different lengths of time. Since my outcome variable is measured when all children in the sample are fourteen, I do not control for age, but do include the other explanatory variables from the Currie and Thomas specification. I include year of *Kindergarten* attendance dummies in all specifications since Figure 4 shows a large increase in attendance over time, which is something I need to account for in this estimation. I also include information on the percentage of children attending either *Gymnasium* or *Realschule* in a given child's state when she was 14 since place constraints influence the probability of whether or not a child attends one of these schools.

I begin my analysis with OLS estimation of the linear probability model for the siblings sub-sample in Table 11 in order to establish a baseline estimate against which I can compare the fixed effects estimation results. In order for OLS to produce a consistent estimate of the parameter β_3, the coefficient on years of early education in the LPM in Equation 2, the unobserved family-specific time-invariant effect must be linearly independent of the regressors, i.e. $E(\eta_j \mid x_i, z_j, yearskind_i) = 0$. Fixed effects estimation will allow me to obtain a consistent estimate of β_3 since it allows me to difference out η_j from the LPM.

Following estimation using both methods, I will use a statistical test based on the work of Mundlak (1978) to determine if the OLS estimates are consistent. The Mundlak test is similar to a Hausman (1978) test, which would also allow me to compare random effects estimates with fixed effects estimates, but more robust than a Hausman test when we have heteroskedasticity in the error term. Since the LPM introduces heteroskedasticity into the model, which I account for with cluster-robust standard errors, the Mundlak test is preferable to the Hausman test.

The OLS estimates of the effect of years of early education on the likelihood of being in a *Gymnasium* or *Realschule* at age fourteen on the siblings-only sample are similar to the OLS estimates on the full sample in Table 9, although the sample size is significantly reduced due to missing values. Here I only find similar evidence of heterogeneous treatment effects for children from disadvantaged backgrounds, i.e. less educated families. Although interesting, these results do not capture any of the unobserved heterogeneity between children or families.

I now turn to family fixed effects estimation to this siblings-only sample. The results in Table 12 show that moving from OLS to FFE estimation does not change the significance of the coefficients on years of early education attendance, but it does change the sign and magnitude. In these specifications, all of the coefficients on years of early education are small and negative. The results of replicating the Currie and Thomas methodology are shown in Column (5) of Table 12, although these results do not differ much from my own approach. In Column (6) of Table 12, I include all child-specific variables that vary enough across siblings to be included in the family fixed effects specification, which decreases the sample size due to missing values in some of these variables. The difference in coefficient size and sign seems to indicate that there is a difference in using OLS versus FFE and that perhaps the OLS regressions are misspecified.

I test the FFE identification strategy using the Mundlak test and the following condition

$$E(\eta_j \mid x_i, z_j, yearskind_i) = 0 \tag{7}$$

which if it holds means the OLS estimates are consistent and there is no selection on family characteristics. I use the Mundlak test instead of the more standard Hausman (1978) test because of the heteroskedasticity introduced into the model by the LPM. In the presence of heteroskedasticity or serial correlation, the Hausman test statistic has a non-standard limiting distribution (Wooldridge, 2002), which is why I need to use a robust version of the Hausman test, such as the

Table 11: Initial OLS Estimates on Siblings-Only Sample

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14
everkind	-0.033							
	(0.111)							
yearskind		0.017			0.026	0.023	0.146*	0.007
		(0.023)			(0.031)	(0.027)	(0.079)	(0.027)
1.yearskind			-0.016					
			(0.103)					
2.yearskind			-0.004					
			(0.098)					
3.yearskind			0.004					
			(0.102)					
4.yearskind			0.072					
			(0.112)					
kindage3				0.120**				
				(0.060)				
kindage4				0.025				
				(0.043)				
kindage5				-0.060				
				(0.051)				
kindage6				0.003				
				(0.066)				
yearskind*male					-0.016			
					(0.037)			
yearskind*oldest						-0.014		
						(0.036)		
yearskind*maxhhedu							-0.011*	
							(0.004)	
yearskind*migback								0.026
								(0.044)
male	-0.135***	-0.135***	-0.136***	-0.132***	-0.095	-0.134***	-0.134***	-0.134***
	(0.036)	(0.036)	(0.036)	(0.036)	(0.099)	(0.036)	(0.036)	(0.036)
oldest	0.000	0.004	0.003	0.002	0.006	0.037	0.003	0.007
	(0.042)	(0.043)	(0.043)	(0.043)	(0.043)	(0.097)	(0.043)	(0.043)
maxhhedu	0.040***	0.039***	0.040***	0.039***	0.040***	0.039***	0.067***	0.039***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.017)	(0.009)
migback	-0.093	-0.089	-0.092	-0.098*	-0.091	-0.091	-0.082	-0.150
	(0.057)	(0.057)	(0.058)	(0.058)	(0.058)	(0.057)	(0.058)	(0.117)
fulltime_m3	0.017	0.011	0.007	0.007	0.014	0.011	0.019	0.009
	(0.073)	(0.074)	(0.075)	(0.075)	(0.074)	(0.074)	(0.073)	(0.074)
fulltime_m14	-0.091	-0.093	-0.093	-0.097	-0.093	-0.093	-0.087	-0.091
	(0.059)	(0.059)	(0.060)	(0.060)	(0.059)	(0.059)	(0.059)	(0.059)
fulltime_d3	0.143*	0.145*	0.145*	0.154**	0.147*	0.145*	0.147*	0.148*
	(0.078)	(0.077)	(0.078)	(0.076)	(0.077)	(0.078)	(0.078)	(0.077)
fulltime_d14	0.007	0.009	0.008	0.007	0.008	0.007	0.016	0.009
	(0.062)	(0.062)	(0.063)	(0.062)	(0.062)	(0.062)	(0.063)	(0.062)
east_3	0.032	0.021	-0.014	-0.053	0.006	0.011	0.093	0.046
	(0.139)	(0.138)	(0.144)	(0.144)	(0.147)	(0.139)	(0.143)	(0.146)
east_14	0.103	0.092	0.101	0.102	0.106	0.101	0.030	0.077
	(0.161)	(0.162)	(0.162)	(0.162)	(0.169)	(0.165)	(0.167)	(0.166)
metro_3	0.063	0.061	0.062	0.059	0.061	0.063	0.067	0.061
	(0.067)	(0.066)	(0.067)	(0.066)	(0.066)	(0.067)	(0.067)	(0.067)
metro_14	-0.020	-0.015	-0.016	-0.021	-0.014	-0.016	-0.012	-0.011
	(0.073)	(0.073)	(0.073)	(0.071)	(0.073)	(0.073)	(0.073)	(0.074)
loginc_3	0.035	0.034	0.032	0.036	0.032	0.035	0.034	0.033
	(0.070)	(0.070)	(0.071)	(0.069)	(0.070)	(0.070)	(0.069)	(0.070)
loginc_14	0.135**	0.131**	0.129**	0.127**	0.132**	0.131**	0.122**	0.128**
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
hhsize_3	0.001	0.003	0.002	0.004	0.004	0.003	0.007	0.005
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
hhsize_14	-0.039	-0.036	-0.035	-0.037	-0.036	-0.034	-0.039	-0.035
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)
ssratio_14	-0.046	-0.102	-0.120	-0.145	-0.102	-0.099	-0.063	-0.100
	(0.583)	(0.583)	(0.584)	(0.581)	(0.584)	(0.585)	(0.583)	(0.582)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-1.188*	-1.216**	-1.139*	-1.120*	-1.232**	-1.244**	-1.513**	-1.177*
	(0.605)	(0.599)	(0.609)	(0.602)	(0.599)	(0.606)	(0.628)	(0.609)
Observations	539	539	539	539	539	539	539	539
R-squared	0.225	0.226	0.227	0.233	0.226	0.226	0.229	0.226

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: The results in this table do not differ greatly if estimated using a binary probit

Table 12: Family Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	FE	FE	FE	FE	FE	FE
yearskind	-0.047	-0.046	-0.046	-0.051	-0.043	-0.059
	(0.031)	(0.031)	(0.031)	(0.035)	(0.034)	(0.044)
male		-0.107**			-0.084*	-0.070
		(0.048)			(0.051)	(0.062)
oldest			0.013		-0.013	-0.044
			(0.060)		(0.066)	(0.134)
fulltime_m3						0.137
						(0.216)
fulltime_m14						-0.088
						(0.160)
fulltime_d3						0.116
						(0.147)
fulltime_d14						-0.229
						(0.199)
metro_3						-0.050
						(0.316)
metro_14						-0.519
						(0.683)
loginc_3						-0.003
						(0.124)
loginc_14						0.154
						(0.168)
hhsize_3						0.118
						(0.115)
hhsize_14						-0.063
						(0.140)
ssratio_14						0.022
						(1.645)
hinckindavg				0.140	0.128	
				(0.124)	(0.134)	
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.723***	0.764***	0.713***	-0.209	-0.221	-0.358
	(0.120)	(0.124)	(0.127)	(0.940)	(1.031)	(1.670)
Observations	829	829	829	769	769	552
R-squared	0.817	0.823	0.817	0.833	0.837	0.891

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

method proposed by Mundlak.

As Quinn (2010) points out, if I define the time-invariant family characteristic as being linearly related to the household average of the child-specific regressors,

$$\eta_j \equiv \alpha\overline{yearskind}_j + \gamma\overline{x}_j \tag{8}$$

and include these means in a linear regression along with my other child-specific covariates, my results will replicate the results of the standard fixed effects estimation. Defining the time-invariant family characteristic in this way allows me to test the following null hypothesis using a joint test on the so-called "Mundlak means":

$$H_0: \quad \alpha = 0, \gamma = 0 \tag{9}$$

The results of this test will tell me if selection based on family time-invariant characteristics matters and whether or not I need to use a fixed effects estimation approach instead of OLS. I follow this methodology only for the Currie and Thomas specification, which includes years of *Kindergarten* attendance, gender, a dummy for being the first born, and household income during *Kindergarten*. The averages in Equation (8) are calculated at the household level "j" for each child "i." In Table 13, I have included the fixed effects estimation following the Mundlak procedure. Columns (1)-(3) of Table 13, denoted "C&T," follow the Currie and Thomas specification discussed earlier, with Columns (2) and (3) looking at heterogeneous treatment effects. I look only at the Currie and Thomas specification since there is enough variation between the siblings included to use these child-specific variables.

Table 13: Mundlak (1978) Fixed Effects

	(1)	(2)	(3)
	C&T	C&T	C&T
VARIABLES	gymreal14	gymreal14	gymreal14
yearskind	-0.043*	-0.070**	-0.063**
	(0.023)	(0.032)	(0.028)
male	-0.084**	-0.196**	-0.084**
	(0.034)	(0.094)	(0.033)
yearskind*male		0.047	
		(0.033)	
hinckindavg	0.140	0.133	0.143
	(0.086)	(0.088)	(0.087)
oldest	-0.013	-0.014	-0.113
	(0.033)	(0.033)	(0.090)
yearskind*oldest			0.041
			(0.033)
Year dummies	Yes	Yes	Yes
Constant	-1.986***	-1.999***	-2.080***
	(0.359)	(0.362)	(0.368)
Observations	769	769	769
R-squared	0.162	0.163	0.165

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The Mundlak test of the coefficients on the household average terms in Column (1) of Table 13 has a p-value of 0.000, which leads to rejection of the null in Equation (9), meaning that family unobserved time-invariant characteristics are not linearly independent of the regressors and Condition (7) does not hold. This implies that the OLS estimates are inconsistent since selection based on time-invariant family characteristics is occurring.

In this case, using a fixed effects estimator, based on the Mundlak methodology, will produce consistent estimates of the returns to early education on later schooling outcomes. It is important to note that the Mundlak test only provides an answer to one dimension of the suitability of FFE over OLS; the OLS results may still be more robust. The results of this alternative fixed effects estimation in Column (1) of Table 13 show a small, negative coefficient on years of *Kindergarten* attendance, which is statistically significant at the ten percent level. A negative coefficient on the years of early education variable here might arise if the early education programmes the children in this sample attend are not pedagogical or are very low quality. If this is the case, children could potentially be better off at home with their parents in terms of development and later outcomes.

This FFE result differs from the positive, insignificant coefficient for years of early education attendance estimated using OLS on this same sample in Table 11. One possible explanation for this is that children with less educated mothers are less likely to attend an early education programme for longer, either because their mothers do not value education as much or because they do not work and so can stay at home with their children. These same children are also less likely to attend a *Gymnasium* or a *Realschule* because they have a less educated mother. This would cause an upward bias on the OLS estimates. Even though I control for highest parental education in the household in the OLS models, if fathers are consistently the highest educated, then the effect of a lesser educated mother is not being controlled for. The introduction of family fixed effects, however, will eliminate selection based on family characteristics, which might explain why the upward bias on the OLS disappears and the returns to early education on secondary schooling outcomes become negative.

Family fixed effects estimation is not without issues. The OLS estimates may still be more robust than the FFE estimates if there is not enough variation between siblings' years of attendance to obtain robust estimates along this margin. Currie and Thomas also point out that the family fixed effects estimates will only be consistent provided that selection into early education is based purely on family characteristics and not on child-specific characteristics. This is a strong assumption and might not be the case in practice since parents might send a less able child to an early education programme for longer so that the child might be able to catch up, causing the fixed effects estimates to be biased downwards, or they might wish to push the more able child by enrolling her for longer so as to maximise future educational outcomes, in which case the fixed effects estimates would overstate the effect of attendance on secondary schooling outcomes.

It is also possible that time trends or birth order are driving the difference in years of attendance. If children are more likely to attend over time, then siblings born later are more likely to attend for longer. Younger and older siblings may attend for different lengths of time depending on when their mother decides to return to work. There may also be differences in access to early education programmes across time that affect younger and older siblings. Especially during the period considered, when there was an expansion of early childhood education facilities, younger siblings born later may have better access to facilities than their older siblings did. This type of systematic variation in the difference between years of attendance may make FFE estimates less robust.

Unfortunately there is no measure in this data set that would allow me to control for innate ability or intelligence, which means I am unable to determine if selection based on unobserved child-specific characteristics is taking place and if so, in which direction. There are also other types of unobserved child characteristics, such as behaviour or social skills, which might be correlated with ability and also affect selection into early education programmes, for which I also have no controls. Any of these child-specific unobservables could be driving the negative returns to years of early education in this context.

Currie and Thomas identify two additional reasons for a potential bias towards zero: measurement error and spillover effects of attending within the household. The results produced by the family fixed effects estimation on this siblings-only sample should not be extrapolated to the general population since within family factors might also be driving them. Spillover effects from the siblings who go to an early education programme for longer might mean that the siblings who go for less time learn the same amount in the home and therefore depress the effect of going for more years.

Of course, having a child attend an early education programme changes the dynamics within the household. Early education programmes also have the function of providing childcare and this traditionally allows the mother to increase her labour supply. Not only can this change the income and consumption patterns of a household, but it can also change the mother's time use dynamics. These intra-family changes might affect secondary schooling outcomes through the years of early education variable in such a way that they are driving the effect to be negative.

In the context of the literature previously discussed in this chapter, the lack of a statistically significant, positive effect of early education on secondary schooling placement is actually not surprising. These programmes are heterogeneous and target the general population. They are not the high quality, high intensity interventions studied in much of the literature where positive effects on academic outcomes were found. Dustmann et al. (2012) also find no effect of early education on school readiness for their native German population. Cascio and Schanzenbach (2013) find no effect of universal early education on academic outcomes after grade 8, which is the same age cut-off used here. Blanden et al. (2015) also find evidence of a fade-out of positive effects after age eleven. Given the institutional context, a finding of a small, negative average effect across this sample is reasonable, which is why I will now turn my attention to heterogeneous treatment effects and non-linearities.

2.6.3 Heterogeneous Treatment Effects

The idea of a treatment affecting individuals in heterogeneous ways has become well accepted in the returns to education literature (Heckman et al., 2006). In order to investigate this possibility, I interact the explanatory variable of interest, years of early education attendance, with the variables for male and oldest sibling to see if the effect of years of early education attendance on secondary schooling outcomes is fundamentally different for boys or girls and younger or older siblings. Since child psychologists emphasise the role of gender and birth order in developmental issues, it seems plausible that the effects of years of early education vary for children along the lines of these classifications.

The results of adding these interaction terms to the fixed effects specification following the Currie and Thomas specification may be seen in Columns (2) and (3) of Table 13. Both of the coefficients on these interaction terms are small and positive, but not statistically significant. The standard errors on these estimates, however, are too large to rule out heterogeneous treatment effects along these child specific characteristics.

There are other variables of interest that might explain heterogeneous effects of early education, such as whether or not the child comes from an immigrant family or the education level of the parents; however, because such family-specific characteristics are controlled for by the family fixed effects specification, this cannot be investigated in the same manner as the child-specific ones. It is possible, however, to estimate a fixed effects regression on the immigrant background sub-sample only. Coming from an immigrant background might alter the way in which families select into early education if immigrants are less familiar with the German education system and lack complete information about their options. Looking only at immigrant families with two or more children significantly decreases my sample size, which limits my ability to comment on the significance of these results.

The results of the OLS estimates when the siblings-only sample is restricted to children from immigrant backgrounds may be seen in Columns (1) and (2) of Table 14. In Columns (3)-(5) of this table I include the family fixed effects estimation results. The initial OLS estimates on years of early education attendance are small and positive, with a magnitude slightly larger than the OLS estimates on the full siblings-only sample; however, once I add in covariates, the coefficient on the years of early education variable becomes much smaller. This is similar to the OLS regression on the full siblings-only sample. In this specification, however, the impact of family income and parental education is larger. Part of this could be driven by an underlying difference between children from immigrant and non-immigrant families in this sample. As shown in Table 5 in the descriptive statistics, the children from immigrant families have on average less educated parents and lower log household income.

Table 14: OLS and FE for the Immigrant Siblings Sample

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	FE	FE	FE
VARIABLES	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14
yearskind	0.050	0.004	-0.087**	-0.125*	-0.091
	(0.034)	(0.044)	(0.043)	(0.070)	(0.055)
male		-0.178**	-0.064	-0.185	-0.063
		(0.069)	(0.062)	(0.162)	(0.063)
oldest		-0.152*	-0.155**	-0.163**	-0.167
		(0.087)	(0.073)	(0.074)	(0.178)
yearskind*male				0.062	
				(0.072)	
yearskind*oldest					0.005
					(0.073)
maxhhedu		0.055***			
		(0.019)			
fulltime_m3		-0.091			
		(0.115)			
fulltime_m14		-0.163*			
		(0.098)			
fulltime_d3		-0.144			
		(0.122)			
fulltime_d14		-0.056			
		(0.115)			
east_3		0.068			
		(0.234)			
metro_3		0.114			
		(0.112)			
metro_14		-0.066			
		(0.133)			
loginc_3		-0.043			
		(0.082)			
loginc_14		0.319***			
		(0.112)			
hhsize_3		-0.039			
		(0.027)			
hhsize_14		0.016			
		(0.044)			
ssratio_14		1.177			
		(1.033)			
hinckindavg			0.063	0.074	0.048
			(0.166)	(0.172)	(0.167)
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant	0.363**	-2.389***	-1.359*	-1.374*	-1.476**
	(0.166)	(0.883)	(0.739)	(0.740)	(0.746)
Observations	274	182	255	255	255
R-squared	0.068	0.311	0.136	0.137	0.149

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: OLS and FE for the Lower Education Sample

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	FE	FE	FE
VARIABLES	gymreal14	gymreal14	gymreal14	gymreal14	gymreal14
yearskind	0.039*	0.017	-0.051	-0.047	-0.064*
	(0.022)	(0.030)	(0.031)	(0.042)	(0.039)
male		-0.157***	-0.142***	-0.126	-0.143***
		(0.047)	(0.049)	(0.117)	(0.049)
oldest		-0.012	-0.020	-0.020	-0.084
		(0.052)	(0.042)	(0.042)	(0.119)
yearskind*male				-0.007	
				(0.048)	
yearskind*oldest					0.028
					(0.046)
maxhhedu		0.161***			
		(0.028)			
fulltime_m3		-0.026			
		(0.091)			
fulltime_m14		-0.107			
		(0.075)			
fulltime_d3		0.093			
		(0.091)			
fulltime_d14		0.032			
		(0.072)			
east_3		0.014			
		(0.140)			
metro_3		0.148			
		(0.106)			
metro_14		-0.114			
		(0.114)			
loginc_3		0.026			
		(0.082)			
loginc_14		0.104			
		(0.078)			
hhsize_3		0.013			
		(0.024)			
hhsize_14		-0.050			
		(0.037)			
ssratio_14		-0.593			
		(0.761)			
hinckindavg			0.087	0.086	0.083
			(0.116)	(0.116)	(0.116)
Year dummies	Yes	Yes	Yes	Yes	Yes
Constant	0.432***	-1.979**	-1.365**	-1.362**	-1.430**
	(0.114)	(0.819)	(0.568)	(0.574)	(0.575)
Observations	533	365	491	491	491
R-squared	0.064	0.249	0.128	0.128	0.130

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Once I use a family fixed effects approach, the sign on the years of early education variable changes to become negative and it also becomes significant. This estimate of -8.7 percent is larger than the family fixed effects estimate produced on the entire siblings-only sample in Table 13, but has the same sign. A Mundlak test of the means included in this fixed effects specification yields a p-value of 0.037, meaning I reject the null hypothesis that the OLS estimates are consistent at the 5 percent significance level. The rejection of the null hypothesis in this case, as in the full siblings-only case, implies that selection based on family unobservables matters and the OLS estimates are inconsistent. These results indicate that there is no fundamental difference in how years of early education affect the children with an immigrant background in this data set once I control for selection based on family characteristics. This result differs from Dustmann et al. (2012), who found positive benefits from early education for immigrants, but not for native German pupils. Due to my small sample size, however, the estimates obtained here are not very robust.

In Columns (4) and (5) of Table 14 I explore the possibility of heterogeneous treatment effects for the immigrant, sibling sub-sample. Again, as in the full sibling sample, the coefficients on the interaction terms are positive, but not statistically significant. The standard errors are too large to assess whether boys and girls and younger and older siblings benefit in different ways from early education.

I also create a sub-sample of children from less educated households. Here I define "less educated" to mean less than twelve years of education and training, which excludes children whose parents have attended a technical school, a higher technical college, or a university. Table 15 presents the OLS and FFE effect results from this analysis. These results do not differ significantly from the results on the full sibling sample and again, the standard errors are quite large, limiting my ability to comment on heterogeneous treatment effects.

Unfortunately I am not able to explore heterogeneous treatment effects due to child ability due to the lack of available measures. As discussed earlier, this potential channel for determining how early education impacts later schooling outcomes could be crucial in explaining the overall negative or zero impact of years of early education on secondary schooling placement. Whilst the average impact might be zero, decomposing this effect by ability could prove useful in better understanding these mechanisms. I leave this question to future research using other more suitable data.

2.6.4 Non-linear Treatment Effects

In all of the specifications of the linear probability model previously presented, I assumed that years of early education entered the model linearly. This means that every year of early education must have the same impact on secondary schooling outcomes, something that might not necessarily be true. In order to test this assumption of linear effects, I estimate the same family fixed effects model, but allow years of early education attended to enter the model differently. There are two ways I will introduce non-linear treatment effects. Firstly, by including a higher order term of the years of early education variable and secondly, by including a dummy variable for each year of early education attended.

Columns (1)-(2) of Table 16 show the results of introducing a quadratic term for the years of early education variable. These results show no evidence to support a quadratic function better fitting the data than the linear model.

In Columns (3) and (4), I introduce dummy variables to look at the impact of each year separately. None of the coefficients on the dummy variables in Column (3) are statistically significant, and only the coefficient on the dummy variable for attending an early education programme for three years is negative. These results do not indicate a non-linear effect of attendance for different years. In Column (4), the dummy variables account for *Kindergarten* attendance at various ages instead of total number of years. This allows me to differentiate potential effects based on child development. The coefficient on the dummy variable for attendance at age five is the only coefficient of these dummies that is statistically significant; it is also negative and rather large. This is the age at which many children would start *Kindergarten* in Germany, even though according to the descriptive statistics presented in Figure 1 show that on average more children attend *Kindergarten* at age six. This could have to do with the fact that most children in Germany start school in August once they are aged six and therefore would most likely attend one year of early education at which point they could already be six years old. It is still unclear why starting *Kindergarten* at this late point would lead to a decrease in the probability of attending a better secondary school. Without child level ability measures, it is difficult to further explore this mechanism.

2.7 Quasi-experimental Analysis

In the previous section of this chapter, a fixed effects approach was constructed and implemented in order to control for the selection into education bias. An alternative way to deal with non-randomised assignment to treatment is to find an exogenous shock, which essentially changed how people select into treatment, and use this shock to calculate the effect of the policy over time as well as the Local Average Treatment Effect (LATE) for the people induced into treatment by this shock.

The following analysis exploits a federal law change (§24 in the *Sozialgesetzbuch* [SGB]) in 1996, which guaranteed all children a place in an early education programme, a *Kindergarten*, from age three until they enrolled in primary school. Quality of publicly provided places tends to be high and rather homogeneous; for example, there are regulations concerning the ratio of teachers to pupils and early education places in Germany are highly subsidised, such that parents on average

Table 16: Non-linear Treatment Effects

VARIABLES	(1) gymreal14	(2) gymreal14	(3) gymreal14	(4) gymreal14
yearskind	-0.059	-0.061		
	(0.092)	(0.136)		
yearskind_sq	0.003	0.0004		
	(0.019)	(0.029)		
1.yearskind			0.143	
			(0.186)	
2.yearskind			0.051	
			(0.171)	
3.yearskind			-0.091	
			(0.177)	
4.yearskind			0.045	
			(0.208)	
kindage3				0.096
				(0.092)
kindage4				-0.079
				(0.080)
kindage5				-0.221**
				(0.104)
kindage6				0.065
				(0.100)
male		-0.070	-0.076	-0.054
		(0.062)	(0.062)	(0.059)
oldest		-0.044	-0.028	-0.053
		(0.134)	(0.126)	(0.127)
fulltime_m3		0.137	0.104	0.091
		(0.216)	(0.196)	(0.216)
fulltime_m14		-0.088	-0.057	-0.054
		(0.161)	(0.140)	(0.152)
fulltime_d3		0.116	0.068	0.086
		(0.149)	(0.154)	(0.144)
fulltime_d14		-0.229	-0.261	-0.211
		(0.198)	(0.189)	(0.187)
metro_3		-0.050	-0.049	-0.041
		(0.313)	(0.279)	(0.296)
metro_14		-0.520	-0.375	-0.559
		(0.690)	(0.757)	(0.660)
loginc_3		-0.003	0.003	0.024
		(0.124)	(0.116)	(0.133)
loginc_14		0.154	0.138	0.146
		(0.168)	(0.163)	(0.171)
hhsize_3		0.118	0.105	0.122
		(0.115)	(0.104)	(0.114)
hhsize_14		-0.062	-0.050	-0.071
		(0.144)	(0.140)	(0.135)
ssratio_14		0.019	0.051	-0.170
		(1.607)	(1.628)	(1.648)
Year dummies	Yes	Yes	Yes	Yes
Constant	0.735***	-0.356	-0.412	-0.408
	(0.154)	(1.643)	(1.645)	(1.753)
Observations	829	552	552	552
R-squared	0.817	0.891	0.898	0.899

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

end up paying 10 percent of the cost (Dustmann et al., 2012). This combination of high quality and low price led to high demand and a shortage of places with long waiting lists before this law change in 1996 (Dustmann et al., 2012). When the law change came into effect in 1996, many new facilities had to be built to accommodate the excess demand. Parents had, and still have, the right to take their county to court if the local government fails to provide a place for their child.

This federal law change meant that after 1996, there were more *Kindergarten* places for children aged three onwards than before, which serves as the exogenous shock or variation in my analysis because the individuals had no control over this. Figure 4 in the descriptive statistics portion of this chapter showed that post-1996, there was an increase in the percentage of children under age six in my data set who attended an early education programme. In this portion of the chapter I focus on children who attended for two or fewer years and children who attended for more than two years since attendance at the extensive margin was fairly universal before and after the policy change. I choose to look at two or fewer years of attendance since on average most children in my sample attend for approximately two years. This allows me to look at below and above average attendance.

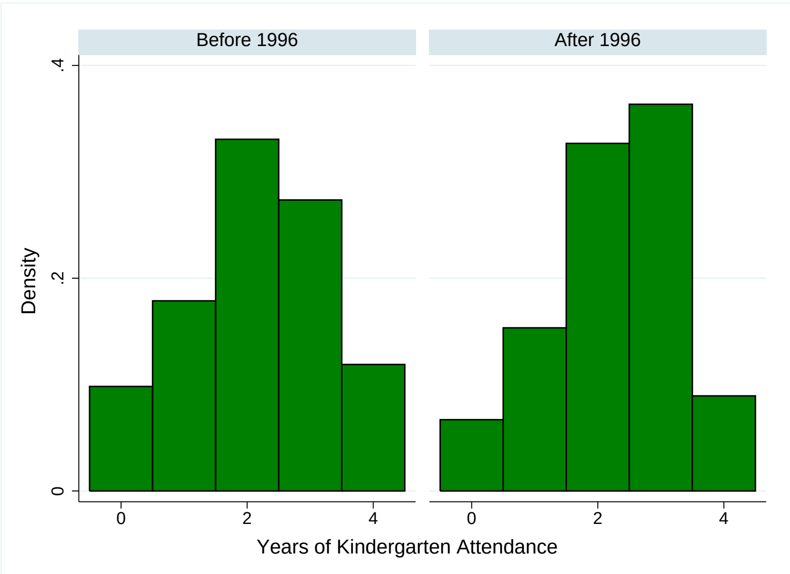


Figure 5: Years of Early Education Attendance for Children Before and After 1996

Figure 5 shows the distributions of years of attendance before and after 1996 in West Germany only.[7] As these histograms show, there is more density to the right of two after 1996, indicating an increase in years of attendance post-1996.

In West Germany, there is regional variation as to how this policy was implemented because of the building of new facilities. Whilst the policy under review in this chapter was enacted on the federal level for all states on January 1, 1996, not all states had the capacity to provide the

[7]In this section of the chapter, I look only at West German states because exploiting regional differences between East and West would most likely violate the common time trend assumption.

necessary number of spaces to accommodate all children by this date. Most notably, the prime minister of North Rhine-Westphalia (NRW) announced that his state would not be able to provide the necessary number of places before 1998, as in 1996 it was still lacking some 250,000 spaces (Georgii, 1996). NRW was the only state to officially announce their capacity constraints. This statement prompted the federal government to allow states an official "transition period" until 1998, by which point they needed to provide enough spaces or risk being sued by the parents of those children who did not receive a space. It is possible that other states were also capacity constrained during this period, but due to limited data and official statements from these states, I am not able to verify this.

Despite the transition period, I still use 1996 as the cut-off point for my before and after since it was the official law change and parents were informed of it. Because of the transition period, however, families in NRW experienced a slower treatment as the state was more constrained and that the first two years of treatment serve as more of an intended treatment effect. Using this information, I restrict the analysis to West German states only and compare NRW with the other West German states. Meghir and Palme (2003) similarly examine an education policy reform in Sweden that was carried out in different provinces at different times, which also allows them to exploit both time and geographical variation. Dustmann et al. (2012) look at the same federal law change in Germany, but only in one state. They find that increased access to early education benefitted immigrant children in terms of school readiness, but not their native German peers.

2.7.1 Difference-in-differences

In this section of the chapter, I exploit the variation in implementation time and geography to assess the impact of attending early education in Germany. I first look at simple difference-in-difference (DID) of pre- and post-1996 in both geographic regions separately in Tables 17 and 18 and then move on to a triple difference (DDD) comparing both geographic regions. These tables show the percent of children who attended a *Gymnasium* or *Realschule* and also attended an early education programme for less than two years or more than two years before and after 1996. Each of these four means and their associated standard errors are presented in the tables. For example, 58 percent of my sample who attended an early education programme for more than two years before 1996 went to a *Gymnasium* or *Realschule*. Once these four means have been calculated, I am able to calculate the differences along each dimension, which are obtained by simply subtracting one mean from the other, and then the DID, shown at the bottom of the far right column, which is the subtraction of one difference from another.

The DID in Table 17 is negative and small, indicating that the expansion of early education pre- and post-1996 did not increase the probability of attending a better secondary school for those children who attended for more than two years. The DID in Table 18 is also negative, but

Table 17: Difference-in-Differences: The Rest of West Germany (Not including NRW)

	Before 1996	**After 1996**	**Difference**
Children Who Attended for More than Two Years	0.582	0.713	0.131
	(0.021)	(0.040)	(0.048)
N	541	129	670
Children Who Attended for Two Years or Less	0.554	0.706	0.152
	(0.017)	(0.030)	(0.036)
N	870	231	1,101
Difference	0.028	0.007	-0.021
	(0.027)	(0.050)	(0.060)
Total number of children	1,411	360	1,771

NB: Standard errors in parentheses

Table 18: Difference-in-Differences: North Rhine-Westphalia

	Before 1996	**After 1996**	**Difference**
Children Who Attended for More than Two Years	0.676	0.652	-0.024
	(0.039)	(0.071)	(0.080)
N	148	46	194
Children Who Attended for Two Years or Less	0.571	0.550	-0.021
	(0.024)	(0.056)	(0.061)
N	403	80	483
Difference	0.105	0.102	-0.003
	(0.047)	(0.091)	(0.102)
Total number of children	551	126	677

NB: Standard errors in parentheses

Table 19: DID NRW Sub-sample

VARIABLES	(1) gymreal14	(2) gymreal14
morethan2	0.105**	-0.043
	(0.047)	(0.057)
post96	-0.021	-0.281***
	(0.060)	(0.106)
post96morethan2	-0.003	0.087
	(0.102)	(0.122)
male		-0.085*
		(0.044)
oldest		-0.104*
		(0.054)
maxhhedu		0.042***
		(0.011)
migback		-0.105*
		(0.062)
fulltime_m3		0.095
		(0.106)
fulltime_m14		-0.083
		(0.078)
fulltime_d3		0.117
		(0.088)
fulltime_d14		0.087
		(0.087)
metro_3		-0.108
		(0.100)
metro_14		0.035
		(0.098)
loginc_3		0.103
		(0.078)
loginc_14		0.089
		(0.071)
hhsize_3		-0.052*
		(0.028)
hhsize_14		-0.008
		(0.026)
ssratio_14		0.492
		(1.999)
Constant	0.571***	-1.241
	(0.024)	(0.856)
Observations	677	390
R-squared	0.009	0.255

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

smaller, indicating that the expansion of provision pre- and post-1996 in NRW did not increase the probability of attending a better secondary school for those children who attended early education for more than two years; however, this difference is not statistically significant. The sample size for children living in NRW is smaller than for the rest of West Germany; however, NRW is the most populous state in Germany, which still gives me almost 700 children.

In order to further examine these DID results, I present the DID regressions on the NRW subsample in Table 18. In these regressions, I cluster the robust standard errors at the household level, which is why they differ from the standard errors presented in Table 18. The results in Column (1) show a statistically significant impact of attending early education for more than two years on secondary schooling placement and an insignificant, negative additional effect of attending for more than two years after 1996. Once covariates are added to the specification in Column (2), the effect of going for more than two years disappears and the effect of attending after 1996 becomes negative and statistically significant. The DID term is not statistically different from zero in either column.

I now turn to the possibility of a triple difference (DDD) by directly comparing NRW with the other West German states as an additional source of variation. Hamermesh and Trejo (2000) employ a similar strategy by comparing men's hours worked in California before and after a law change to extend the requirement for overtime to men in other states and then to women. Kellogg and Wolf (2008) compare the effect of daylight savings on energy use by looking at Sydney and its extended daylight savings for the Olympic Games compared to the rest of Australia and then looking at night time use versus mid-day use. Here I compare children who attended *Kindergarten* for more than two years or two or fewer years before 1996 ended or after across the two regions. This produces eight comparison groups:

Group A Children from NRW, who went to *Kindergarten* for more than two years before 1996

Group B Children from NRW, who went to *Kindergarten* for more than two years after 1996

Group C Children from West Germany, but not NRW, who went to *Kindergarten* for more than two years before 1996

Group D Children from West Germany, but not NRW, who went to *Kindergarten* for more than two years after 1996

Group E Children from NRW, who attended *Kindergarten* for two or fewer years before 1996

Group F Children from NRW, who attended *Kindergarten* for two or fewer years after 1996

Group G Children from West Germany, but not NRW, who attended *Kindergarten* for two or fewer years before 1996

Table 20: DDD Groups' Secondary Schooling Conditional Means

Group	**Conditional Mean**	**SE**	**N**
Group A	0.676	0.039	148
Group B	0.652	0.071	46
Group C	0.582	0.021	541
Group D	0.713	0.040	129
Group E	0.571	0.025	403
Group F	0.550	0.056	80
Group G	0.554	0.017	870
Group H	0.706	0.030	231

Group H Children from West Germany, but not NRW, who attended *Kindergarten* for two or fewer years after 1996

Each of these comparison groups essentially receives a variation in dosage of the treatment, early education, based on where they lived and when they attended *Kindergarten*. The secondary schooling mean outcomes for each group may be seen in Table 20.

I calculate the DDD using the mean values of the secondary schooling outcome variable for each of these different groups presented in Table 20 as follows:

$$\begin{aligned}
DDD &= DD_{Affected} - DD_{Unaffected} \\
DDD &= DD_{NonNRW} - DD_{NRW} \\
DD_{NRW} &= E\left(Y_{kt} \mid k=1,\, t=1,\, NRW\right) - E\left(Y_{kt} \mid k=1,\, t=0,\, NRW\right) \\
&- \ E\left(Y_{kt} \mid k=0,\, t=1,\, NRW\right) - E\left(Y_{kt} \mid k=0,\, t=0,\, NRW\right) \\
DD_{NonNRW} &= E\left(Y_{kt} \mid k=1,\, t=1,\, NonNRW\right) - E\left(Y_{kt} \mid k=1,\, t=0,\, NonNRW\right) \\
&- \ E\left(Y_{kt} \mid k=0,\, t=1,\, NonNRW\right) - E\left(Y_{kt} \mid k=0,\, t=0,\, NonNRW\right) \\
DDD &= [[mean(Group\, D) - mean(Group\, C)] - [mean(Group\, H) - mean(Group\, G)] \\
&- \ [[mean(Group\, B) - mean(Group\, A)] - [mean(Group\, F) - mean(Group\, E)] \\
&= (-0.026) + (0.041) \\
&= -0.033
\end{aligned}$$

Here k is the binary dummy variable for attending more than two years ($k = 1$) or two or fewer years ($k = 0$) and t is the binary dummy variable for pre-1996 ($t = 0$) or post-1996 ($t = 1$). This negative result indicates that children who attended more than two years post-1998 as a result of expanded access in West German states (the non-NRW states) saw a decrease in their probability of attending a better secondary school at age fourteen. This could be related to the previously mentioned constraints on secondary schooling places. Although access to early education expanded during this period, the number of *Gymnasium* or *Realschule* places may not have expanded at the same rate meaning that the effectiveness of the treatment is limited.

The same result may be obtained through running the following regression with interaction terms:

$$\begin{aligned} E_i &= \beta_0 + \beta_1 nonNRW_i + \beta_2 Morethan2_i + \beta_3 Post96_i \\ &+ \beta_4 nonNRW * Morethan2_i + \beta_5 Post96 * nonNRW_i \\ &+ \beta_6 Post96 * Morethan2_i + \beta_7 Post96 * nonNRW * Morethan2_i + u_i \end{aligned} \tag{10}$$

The coefficient on the triple interaction term, β_7, is the DDD in the regression in Table 21.

Overall these DID and DDD results do not indicate that there has been a statistically significant effect of the increased access to early education pre- and post-1996 for children who attended for more than two years.

2.7.2 Instrumental Variables

In this section of the chapter, I attempt to control for the endogeneity in my variable of interest by implementing an instrumental variable (IV) strategy. I exploit the variation in provision of *Kindergarten* places on a regional basis and use the dummy variable for North Rhine-Westphalia as an instrument for years of early education attendance. An instrumental variable strategy may be implemented in the case of an endogenous regressor, in this case years of early education attendance, where the endogenous regressor is correlated with the error term. These unobservable components in the error term could include child ability or how much a family values education, for which I do not have measures. This type of endogeneity will lead to inconsistent estimates of the parameter of interest using OLS (Cameron and Trivedi, 2005). In order to obtain consistent estimates, I employ an instrumental variable strategy and two stage least squares estimation. This identification strategy requires the instrument be uncorrelated with the unobserved error term of the regression and have a non-zero covariance with the endogenous regressor (Wooldrige, 2002). The assumption that the instrument is uncorrelated with the error term implies that the instrument should not be a regressor in the model for the outcome variable. The second point implies that there must be an association between the instrumental variable and the variable for which I instrument.

In this case, being from NRW should be correlated with years of early education provision since the state was constrained and unable to offer as many children a *Kindergarten* place in comparison to other West German states, but it should not be correlated with the probability of attending a *Gymnasium* or *Realschule*. As previously discussed, children from NRW had to wait longer to get access to increased early education provision because of the acute shortage of places in that state and their families could not influence this. The lack of correlation between the instrument and the outcome variable proves more difficult to assess. It requires the assumption that being from NRW

Table 21: DDD NRW vs. Non-NRW West

VARIABLES	(1) gymreal14	(2) gymreal14
wnnrw	-0.021	-0.074*
	(0.035)	(0.040)
morethan2	0.095**	-0.035
	(0.047)	(0.053)
post96	-0.024	-0.243***
	(0.063)	(0.081)
wnnrwmorethan2	-0.062	0.040
	(0.056)	(0.063)
post96wnnrw	0.177**	0.318***
	(0.073)	(0.095)
post96morethan2	0.007	0.089
	(0.100)	(0.120)
post96wnnrwmorethan2	-0.033	-0.038
	(0.116)	(0.140)
male		-0.096***
		(0.024)
oldest		-0.006
		(0.026)
maxhhedu		0.047***
		(0.005)
migback		-0.034
		(0.032)
fulltime_m3		0.033
		(0.048)
fulltime_m14		-0.080**
		(0.039)
fulltime_d3		0.069
		(0.047)
fulltime_d14		0.120**
		(0.047)
east_3		0.475***
		(0.114)
metro_3		0.004
		(0.052)
metro_14		-0.001
		(0.054)
loginc_3		0.048
		(0.037)
loginc_14		0.113***
		(0.039)
hhsize_3		-0.007
		(0.013)
hhsize_14		-0.005
		(0.016)
ssratio_14		0.213
		(0.331)
Constant	0.574***	-1.266***
	(0.028)	(0.349)
Observations	2,448	1,383
R-squared	0.013	0.185

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 22: West German Instrumental Variable Regressions

VARIABLES	(1) gymreal14	(2) gymreal14
morethan2	-0.050	-0.046
	(0.342)	(0.256)
male		-0.097***
		(0.024)
oldest		-0.002
		(0.029)
maxhhedu		0.046***
		(0.006)
migback		-0.042
		(0.035)
fulltime_m3		0.024
		(0.059)
fulltime_m14		-0.068
		(0.043)
fulltime_d3		0.058
		(0.048)
fulltime_d14		0.118**
		(0.047)
metro_3		-0.004
		(0.052)
metro_14		0.011
		(0.054)
loginc_3		0.047
		(0.038)
loginc_14		0.113***
		(0.042)
hhsize_3		-0.007
		(0.015)
hhsize_14		-0.007
		(0.016)
ssratio_14		0.287
		(0.390)
Year dummies	Yes	Yes
Constant	0.559***	-1.340***
	(0.153)	(0.382)
Observations	2,448	1,383
R-squared	0.015	0.177
F-statistic	12.70	13.36

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

does not change the probability of attending a *Gymnasium* or *Realschule* as compared to the rest of Germany. In practice this would require that the school systems across the German states are relatively standardised and that place availability as well as demand follow the same trends.

The results of these IV regressions may be seen in Table 22. In Columns (1) and (2), I instrument for the dummy variable of whether or not a child attended an early education programme for more than two years using the dummy variable for being from NRW using two stage least squares. Again, I focus on the variable of attendance for more than two years since this means attending for more years than the sample average. In neither case, does instrumenting give me a statistically significant impact of attending for more than two years on secondary schooling outcomes. I report the F-statistics from the first stage of estimation, which tells us if the coefficients in the first stage regression of the endogenous regressor on the instrument are zero. Both of the F-statistics for these regressions are greater than ten, the rule of thumb critical value established by Staiger and Stock (1997), which means I can reject the null hypothesis. This indicates that weak instruments are not an issue in this just-identified context; however, this does not confirm the exogeneity of my instrument.

Bound et al. (1995) point out the difficulty in choosing valid instrumental variables. In order for NRW to be a valid instrument, it can have no direct effect on secondary schooling outcomes; the only association between the instrument and the outcome must come through the relationship between the instrument and the endogenous regressor. The evidence on attendance at secondary school types contradicts this assumption. As Figure 6 shows, there were differences in *Gymnasium* and *Realschule* attendance in NRW versus the rest of West Germany until approximately 2005. The children affected by the federal law change in 1996, would have started secondary school in the early 2000s. This implies that children from NRW would have had a higher probability of attending a *Gymnasium* or a *Realschule* simply because they were from NRW. This is not a formal test of instrument exogeneity, but might explain why these estimates are biased. The lack of significant, positive results here mirrors the results from the family fixed effects section, but does also not come as a surprise since the instrument may not be not valid.

2.7.3 Issues with the Empirical Strategy

My identification strategy relies on the delay in expansion of provision due to constrained capacity. It is also worth considering how many parents actually took their local government to court in order to secure a *Kindergarten* place. This data would provide an indication of how severe the shortage of places truly was and how long the delay in provision lasted; however, it would not be a perfect measure as many parents might be unwilling to go to court even if they are not provided with a place.

In order to investigate this further, I contacted all fifty-one regional courts in Germany responsible for these type of cases and asked them how many cases parents had brought forward in the post-1996 period. Only the court in the city of Mainz, the capital of the state of Rhineland-Palatinate, was able to provide me with a concrete number of cases (15). Some courts responded that they had had a few cases where the parents sued because their child did not get a place at a specific *Kindergarten*; however, this is not of interest to my research as the children were offered a place at an alternative *Kindergarten*. The other courts either do not collect statistics on this type of information or did not have any such cases. The lack of data on these cases does not imply that parents either did not bring such cases forward or that there was not a shortage, but rather highlights the problem of missing data.

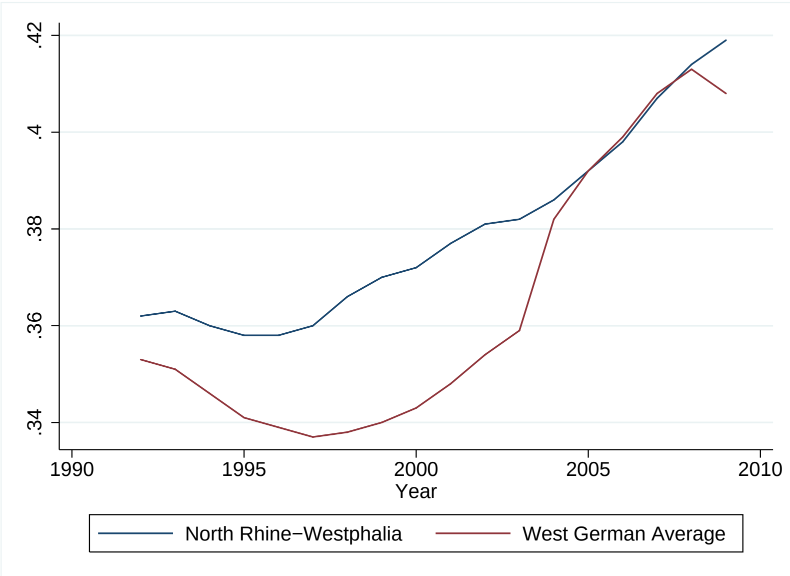


Figure 6: *Gymnasium* and *Realschule* Attendance 1992-2009: NRW vs. West German Average

One potential issue with any DID analysis is the common time trend assumption. Figure 6 suggests that growth in places at *Gymnasium* and *Realschule* in NRW far surpassed the West German average until 2005, when the two equalised. This could cause problems in the accuracy of the DDD estimates since the DDD relies on a common time trend assumption for the two groups in question. I do, however, include a control variable for the percentage of school age pupils in a given state and year in a *Gymnasium* or *Realschule* when I add control variables to the DDD specification. Nevertheless, this limits the strength of the estimation strategy.

2.8 Conclusion

The long-run effects of early education are difficult to measure and explain due to the confounding nature of many other factors, especially family background. In this chapter, I used two different econometric approaches to control for the selection into education based on family characteristics.

My results showed that initial OLS estimates of the returns to early education, which were small and positive, were inconsistent because they ignored this selection based on family characteristics. Once I controlled for selection using a family fixed effects approach on a siblings-only sub-sample of my data set, the returns to years of early education on secondary schooling outcomes became small and negative. Due to data limitations, I was still unable to account for unobserved heterogeneity between children, which could be driving these results.

In the quasi-experimental section of this chapter, I again focused on the problem of selection based on unobservable characteristics, but this time through an exogenous policy change that altered the selection into early education process. Using this exogenous policy change, which expanded access to early education after 1996, I was able to calculate the LATE for the children induced into early education as a result of this policy. One potential constraint preventing children from attending an early education programme could have been the inability to pay the cost. The LATE was only positive in the initial analysis of one German state: NRW. Once I added control variables into this specification, the LATE became negative, and remained statistically insignificant in all other analyses.

The lack of statistically significant results in this chapter and the inability to show a positive effect of *Kindergarten*, even for children from disadvantaged backgrounds, contradicts the dominant narrative in the field perpetuated by Heckman and co-authors (Cunha and Heckman, 2007). The evidence generated by high quality, high cost interventions targeting a specific population has dominated the debate surrounding early education. As discussed in this chapter, these studies find long-term, robust effects of such interventions on learning outcomes and a variety of other measures, especially for children from disadvantaged backgrounds.

High quality, high cost interventions are not the majority of early education programmes, however, and it is crucial to make this distinction clear since most governments do not spend this kind of money on their early childhood programmes. Existing programmes tend to be widespread, low cost, and heterogeneous both in the children they serve and their quality. The results I obtain in this chapter are similar to the U.S.-based studies that explore the expansion of universal early education on the entire population of a state. These studies often find no effect of early education on academic outcomes and some effects on non-cognitive outcomes. It is important that these types of findings also reach policymakers because realistic expectations need to be set about what early childhood education can and cannot do.

The heterogeneity in the quality of early education in Germany is an important issue that I was ultimately unable to address using the data at my disposal. Spiess (2012) notes that the quality of much of the early education in Germany leaves much to be desired, but is not an issue at the centre of the policy debate at the moment; policymakers are more concerned with expanding provision. This may be confounding my results, even after controlling for family-specific characteristics, since variation in treatment also violates the stable unit treatment value assumption.

Currently, Germany's early education landscape is in the midst of a major reform. In August 2013, there was another federal law change to guarantee any child from birth until entry into primary school a place at an early education programme. While many of the state courts I contacted regarding the federal law change in 1996 did not have data on the number of cases following this change, many of the judges predicted that the number of cases following the 2013 law change would be much higher. This is because there is much less capacity to provide early education places for children under the age of three in Germany. An estimated 160,000 places were still missing in Germany as of summer 2012 (Wittrock, 2012). The first parents to take their city to court over the lack of available places have done so in Leipzig and won damages. Interestingly, this case took place in Leipzig, the former East, where there are strong traditions of childcare. This indicates perhaps that since reunification there has been a shift in the former East to resemble the former West in terms of childcare provision.

At the same time, the states of Bavaria, Baden-Württemberg, Saxony, and Thuringia have implemented a childcare policy, known as *"Landeserziehungsgeld"* or "State childcare subsidy," which shifts the burden of childcare or early education away from the state. Under this policy, any family that agrees to not enrol their child in an early education programme, but instead provide it within the home, is given a subsidy to do so (Spiess, 2012). This policy has become an issue of national debate in Germany; the former Federal Minister of Family Affairs, Senior Citizens, Women and Youth, Kristina Schröder, advocated the introduction of a similar subsidy, known as *"Betreuungsgeld,"* at the federal level. Giving a subsidy to parents is less expensive than increasing early education programme capacity by building more facilities and hiring more teachers, especially in light of the August 2013 law change.

This type of childcare subsidy might not have negative consequences for children from high socio-economic status families, but for children from lower socio-economic backgrounds, this might actually hurt their future outcomes. If their parents are financially constrained enough to be induced to take this subsidy and provide childcare on their own, the quality of the children's early education might suffer. This does not take into consideration potential income effects as a result of taking the subsidy.

These policy changes and the debate surrounding them show that both politicians and the public have not reached a consensus on the issue of early education and its returns in Germany. This chapter showed that family and child characteristics play an enormous role in educational decisions and outcomes. The 2013 federal law change, as well as the issue of *Betreuungsgeld* will change the way families in Germany select into early education and provide ample opportunities for future research.

3 School Hours and Female Labour Supply: Quasi-experimental Evidence from Germany

3.1 Introduction

The linkage between childcare and female labour supply has received much attention in labour economics and public policy. This chapter asks if extending the primary school day makes mothers more likely to enter the labour market or extend their hours if already working. It has been shown that the costs of childcare affect how women determine their labour force participation and hours. As the primary carers of children, women often exit the labour market after having children and only re-enter once childcare becomes available and affordable.[8] When preschool childcare is expensive or limited in supply, primary school may be the first opportunity for families to utilise "free" childcare. In many countries, however, it may not be possible because of the length of the primary school day.

Germany, Austria, and Switzerland were the only Western European countries to not have "full length" school days until approximately ten years ago. A typical German school day, at both the primary and secondary level, began around 8am and would end at approximately 1pm. A recent and on-going reform in Germany to increase school hours, known in German as the *Ganztagsschulreform* or "full school day reform" was undertaken in an effort to extend the school day by a few hours until approximately 3pm (Stecher et al. 2008).[9]

In Germany, the debate to increase school hours centred on improving education outcomes, based on the idea that learning is a function of the number of hours spent in the classroom. Impetus to speed up this reform came from the 2001 "PISA Shock," the public reverberation of Germany's rather poor performance on the Programme for International Student Assessment (PISA) tests administered in 2000.

Apart from focusing on educational outcomes, another major reason for tackling the issue of school hours was to increase the labour market participation of mothers. It was argued that not having a full day of schooling was creating a structural barrier to women entering into employment after having children. As labour market statistics for women in Germany show, female employment has increased over the period of the reform.

In this chapter, I ask whether extending the primary school day caused mothers to enter the labour market if they were not working before or extend their working hours if they were already working. Using the randomness of timing in the reform, I am able to exogenously assign treatment, access to a full day school, to women and explore how this impacted their labour market status. I find

[8] This simplifies the decision to re-enter the labour market, ignoring the implicit costs associated with employment such as the penalty many women may pay for having taken time off for maternity decisions.

[9] The details of the reform will be further discussed with the data and as part of the identification strategy.

that women who were not working were able to enter the labour market, but that for women already participating, there is either no effect or a very small negative effect on the hours they work. This result may be explained by the income and substitution effects underpinning a basic model of labour supply.

3.1.1 Female Employment in Germany

Perhaps unsurprisingly, the female employment ratio in Germany is lower than in many European countries. Of course when we talk about Germany today, we refer to a reunified Germany, which brought together two different traditions and attitudes towards female participating in the labour market. Wenzel (2010) points out that these differences, primarily the result of East Germany having had a stronger tradition of women working and as a result more developed childcare options, still persist today. Even in 2002, more than ten years after German reunification, 51.7 percent of mothers in former East Germany were full time employed; this is contrasted with only 16.8 percent of mothers in West Germany (Wenzel, 2010). These differences should be kept in mind when looking at statistics from all of Germany.

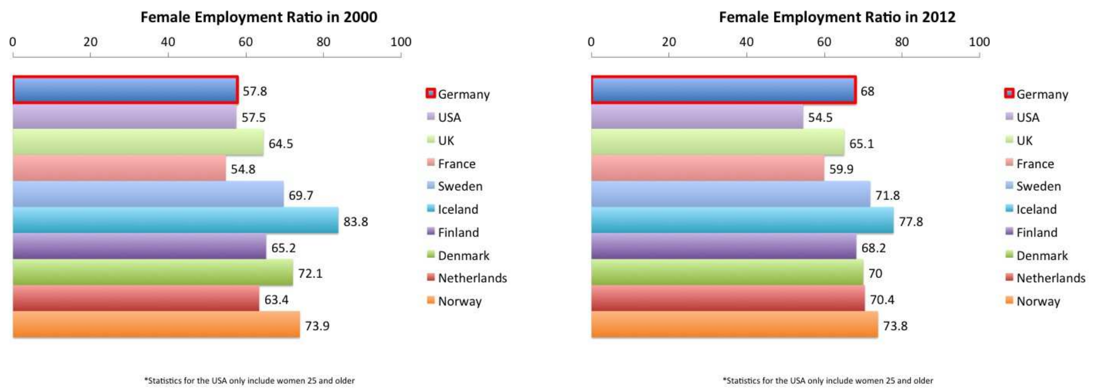


Figure 7: Female Employment Ratio in 2000 and 2012, Source: ILOStat

In 2000, only 57.8 percent of German women 15-64 were employed, either in full time or part time employment (International Labour Organisation Statistics). By 2012, this had increased to 68 percent (ILOStat). Figure 7 shows how Germany compares to the Nordic countries, the Netherlands, France, the United Kingdom, and the United States in 2000 and in 2012[10] in terms of the female employment ratio for women aged 15-64. The data from both years shows that Germany has a lower female employment ratio than most of the Nordic countries, but that the gap has decreased over time.

[10]It should be noted that the data for the United States only contains information on women aged 25 and older due to differences in labour force surveys.

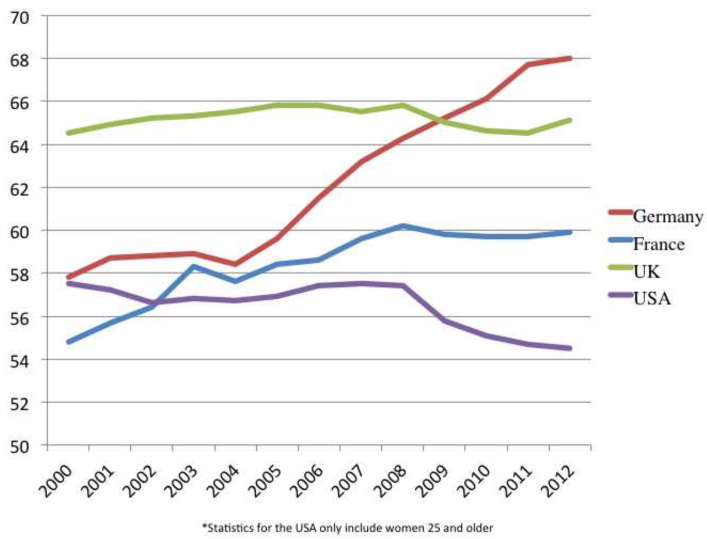


Figure 8: Female Employment Ratio, Source: ILOStat

The evolution of Germany's female employment ratio also differs from France, the United Kingdom, and the United States. Although Germany, France, and the United States start off in similar positions in 2000, Figure 8 shows that Germany's female employment ratio has increased much faster and at a rate that allowed it to overtake the United Kingdom by 2012. Its female employment ratio has increased significantly over this period.

As in many other countries, another characteristic of female employment in Germany is the high incidence of part time work. All of the statistics presented above focused on employment and did not differentiate between part time and full time work, which is also informative for understanding female labour supply in Germany. Of all part time work carried out in Germany, women account for 90 percent of it (ILOStat). Many women in Germany work part time, which may also be seen in Figure 9, a histogram of hours worked by men and women from the data used in this paper. This figure clearly shows that most men in the sample work full time, but that much of the density of hours for women is found towards the left of the distribution. This prevalence of part time work will help explain the results found in this chapter.

3.1.2 Childcare in Germany

The childcare landscape in West Germany is characterised by limited places. Major policy changes in both 1996 and 2013, caused considerable expansion in pre-school places for children aged three to six; however, childcare options for children already enrolled in primary school remain constrained.

Before the school day was extended in Germany, there was the option of after-school care, known as a *Hort*, most often provided by non-profit organisations, but often physically located at the primary school (Riedel, 2005). Parents had to sign their children up for a place at the *Hort* and

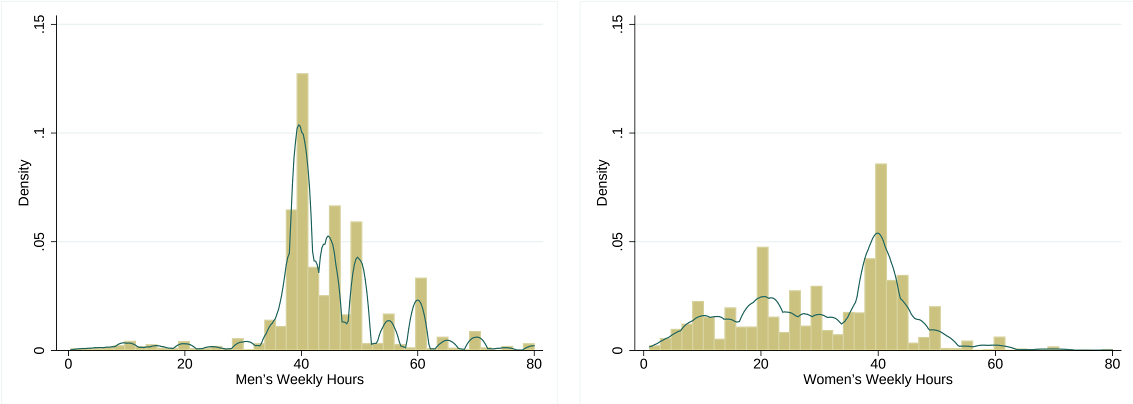


Figure 9: Distribution of Hours Worked For Men and Women

pay for this service, which would often end at 4pm unless they also signed them up for an extended programme (Riedel, 2005). *Hort* still operate at many primary schools in Germany and even at full day schools since working parents may require additional childcare.

Table 23: After School Childcare Place Availability Age 6-10

State	2006 Places	Children	PA	2009 Places	Children	PA	2012 Places	Children	PA
Bavaria	103,613	639,815	16.2	108,121	592,139	18.3	117,255	556,147	21.1
Hesse	56,004	301,950	18.6	58,927	280,988	21.0	59,138	268,690	22.0
Rhineland-Palatinate	29,302	205,163	14.3	24,803	185,738	13.4	23,544	171,342	13.7
Schleswig-Holstein	21,330	148,701	14.3	23,736	136,583	17.4	21,809	125,084	17.4

Source: *Statistisches Bundesamt, Statistik der Kindertagesbetreuung*

Table 23 shows data collected on after-school childcare place availability in the four states of interest in this paper (Bavaria, Hesse, Rhineland-Palatinate, and Schleswig-Holstein). The column "PA" ("Place Availability") shows the number of places available per 100 children in that state. Over time there has been an increase in the number of places available, but still less than 25 percent of children in primary school in these states have access to a childcare place. This is especially low when compared to former East German states, which offer anywhere between 60-80 places per 100 children (Riedel, 2005).

These statistics only reflect institutionally available places and do not reflect the availability of informal childcare options. Nevertheless, the lack of available places serves as a constraint on mothers entering the labour market once their children enter primary school.

3.2 The *Ganztagsschulreform* (Full Day School Reform)

3.2.1 The Reform in the German Educational Landscape

The reform to extend the primary school day in Germany has been an on-going process over the last 10-15 years, born out of the motivation to not only improve educational outcomes, but also to make work and family more manageable for women. The *Ganztagsschulreform* or Full Day School Reform is a reform process to extend the length of the school day at both the primary and secondary schooling levels. In 2006, the *Kultusministerkonferenz*, a regular assembly of all Ministers of Education from the federal states, defined a *Ganztagsschule* as a school that offers at least seven hours of instruction per day for a minimum of three days out of the school week and offers lunch to its pupils (Holtappels et al. 2008). Since education is a federal issue, states agree to have their schools extend the length of the school day according to a timeline they develop. This timeline is based on discussions with the Ministry of Education in each state and the feasibility of transitioning to a full day school. This feasibility is determined in part by the speed at which new teachers may be hired and cafeterias may be built since lunch must now be available, which was not the case under the old system. States also have the flexibility to determine the model of full day schools they wish to implement, something I will discuss later on. I have collected a unique data set on when every primary school in four German states began operating as a full day school (see next section for further discussion of the data and these states) in order to evaluate the impact the reform has had on female labour supply.

The reform to extend the school day has been slow, primarily due to the high costs associated with it, which proves useful for identification. Schools did not switch over all at once; in fact, even within a state, county, or city there is substantial variation over a period of almost 10 years as to when schools switched over from half day institutions to full day ones. This variation arises because of the costs associated with the reform. One of the main costs of extending the school day arises in the necessity of building cafeterias to provide lunch on site. Very few primary schools in Germany would have had a cafeteria or any kitchen facilities on site since all children went home for lunch when school ended. The costs and time lags associated with constructing cafeterias should not be underestimated when assessing the speed of the reform. Much of the funding for the reform has come out of the *Investitionsprogramms "Zukunft Bildung und Betreuung"* (Investment Programme: The Future of Education and Childcare), which committed 4 billion Euros of federal money to the reform during the period 2003-2009 (Rainer et al., 2010). This means that the reform process has been more standardised across states since much of the money comes from a federal source.

Another cost of the reform is the hiring of additional teachers. Teachers in Germany have a special civil servant status, which means the government cannot simply extend their current working hours. Teachers must be converted from part to full time or additional teachers must be hired. The costs associated with hiring new teachers and building new cafeterias are paid by the federal government and state and not by the individual municipality or county, so it seems reasonable to assume that when a school switches is not correlated with other characteristics of the local area in which the school is located, especially since two schools within a relatively homogenous region (e.g. a small city) may differ in their switch-over years. Potentially school switch-over year could be correlated with some school specific characteristic, e.g. the seniority of the principal, however, since I do not look at education outcomes of the pupils, this seems less relevant for mothers' labour supply.

Primary school[11] attendance in Germany is decided solely on proximity to school as there are few private primary schools in Germany and homeschooling is prohibited by law. This allows me to use geographic proximity to a full day school to evaluate the effect of the reform on maternal labour supply. After speaking to people from the Ministries of Education in these four states, it seems that based on their anecdotal evidence, on average less than one percent of families request that their child attend a primary school that is not the school to which they were assigned, i.e. the school closest to their home. They unfortunately do not collect official statistics on this, but if their estimates are accurate, then using closest school as a measure of access seems valid.

Because Germany has a federal system, the education system and the reform process in the four states analysed in this chapter are not identical. The *Kultusministerkonferenz*, a regular assembly of all Ministers of Education from the federal states, ensures, however, that many elements of the education systems are standardised. These four states all have a similar structure to their education system, where children attend primary school from age six until the end of fourth grade, when they are ten year old. At this point, the children are then placed into one of three tracks: the university track secondary school (*Gymnasium*), a higher vocational track secondary school (*Realschule*), and a lower vocational track secondary school (*Hauptschule*) (Robelen, 2005). Because there are no national standardised tests in Germany, it is very difficult, and not encouraged, to compare the education systems across states.

There is one key difference in the reform between the four states, which has to do with whether or not every class at a given school switches to a full day or just a certain percentage of classes switch (in German this is the difference between an *offene Ganztagsschule*, open full day school, and a *gebundene Ganztagsschule*, complete full day school). An *offene Ganztagsschule* might only have one or two classes per grade level that offer the extended school day option and parents would have to choose to sign their child up for this option whereas at a *gebundene Ganztagsschule*, all children automatically receive the longer school day. Regardless of the type of full day school, they still may offer only three days of extended instruction per week; however, in the data collection and contacting of Ministries of Education for this project, it seems that many schools offer the extended instruction every day of the week.

[11] Here primary school aged children are 6 to 10 years old, as secondary school begins in grade 5 in the four states I analyse.

The type of full day school issue, however, may pose a challenge to my analysis. In Bavaria, for example, all primary schools that have switched to the full day are *gebundene Ganztagsschulen*, while in the other states, this has not been the case. Some schools in some of the other states may have switched all classes while others may only have switched one class. The main difference between these two models of switch-over is the cost: switching all classes at the same time means even more teachers must be hired. This accounts for the slower rate of reform in Bavaria as opposed to the other states. For the purposes of the analysis in this chapter, I treat all full day primary schools in the same manner because I assume that having the availability of a full school day is enough for the mother to be treated. In discussions with the Ministries of Education for this chapter, there confirmed that there has not been massive over-subscription or issues of shortage surrounding full day school places, which makes me confident in combining these two types of full day schools in my analysis.

3.2.2 Data on the Reform

Education policy is devolved in Germany, which means that each state has a large amount of autonomy over its education system. The devolution of education policy to the states makes data collection in Germany difficult, but also creates ample regional variation in policy implementation. In this chapter I use a unique, self-collected data set to look at the process of the full day school reform in four West German[12] states: Hesse, Rhineland Palatinate, Schleswig-Holstein, and Bavaria. These four states were chosen because they were the only West German states to have the necessary data centrally collected over the period of interest. This data had never been collected across states in this manner because some states do not collect this information centrally, but rather allow school districts to record this information. Although the data collection process proved arduous, I now have a unique data set never before used for this type of analysis.

For every primary school from the four states in my sample, I have the year in which the reform took effect and the school began operating as a full day facility or a missing value if the school has not yet extended its hours. I observe the first schools operating as full day schools for the 2002-2003 school year; my data continues until the 2012-2013 school year. The switch-over process is still on-going in Germany, and in Bavaria, for example, there are still many primary schools that have not switched over, while in the other states almost half of all primary schools have since transitioned. Figure 10 shows the percentage of total primary schools in each of the four states of interest that have started operating as full day facilities in each year. As may be seen in this figure, less than 50 percent of total primary schools in each state have switched over as of 2012.

In order to analyse the geographic distribution of full day schools and link school level data to

[12] This chapter focuses only on West Germany because of the underlying differences between the West and East German education systems. East Germany already had many schools that offered full school days because women were expected to participate in the labour market; childcare was much more developed in East Germany as a result.

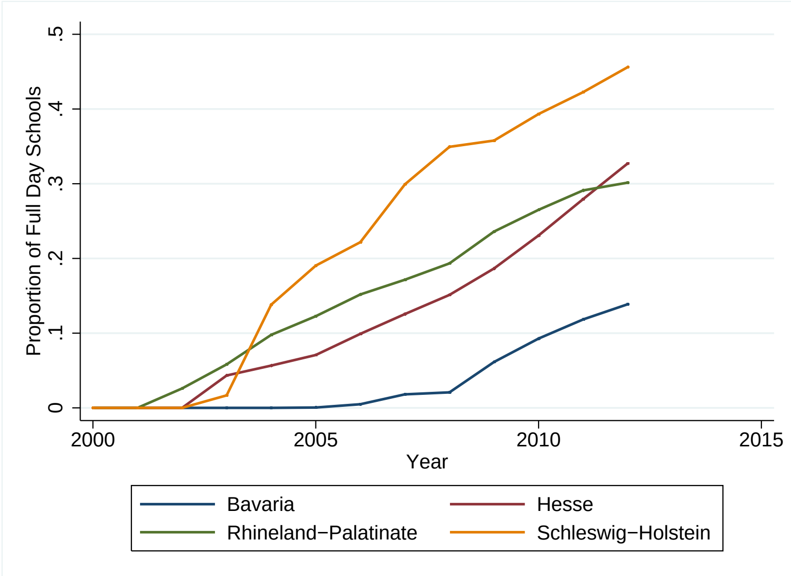
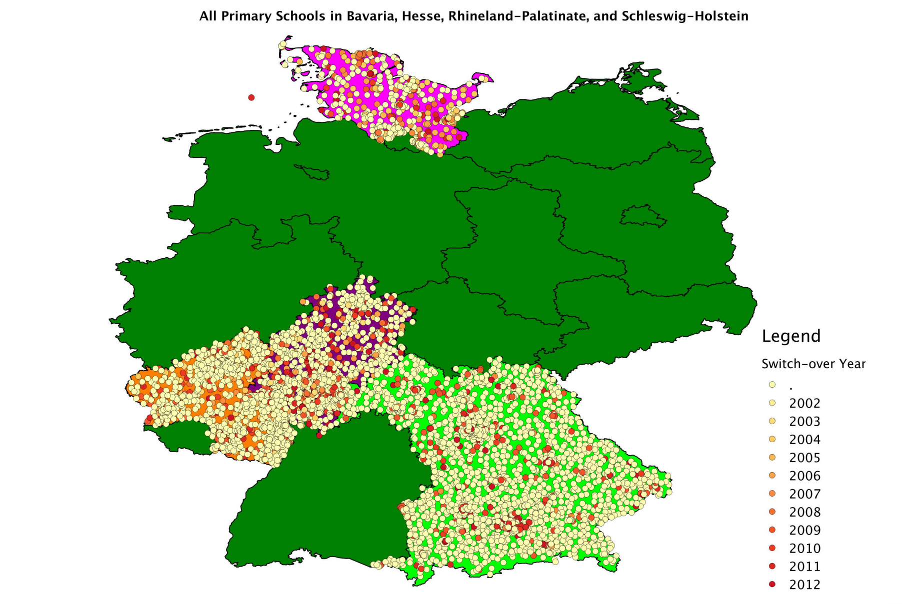
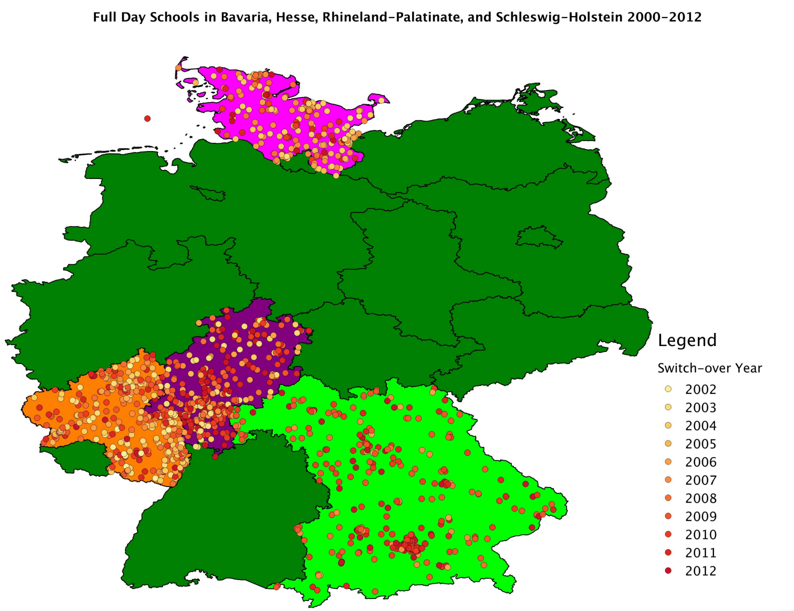


Figure 10: Proportion of Primary Schools Operating as Full Day Facilities By Year

individual level data, I use geographic information system (GIS) software to link a woman to her closest primary school using her geocoded address and geocoded addresses of all primary schools in her state. Unfortunately, the GSOEP does not identify the name of the school a child attends (obviously for women without children it would not do so either), only the type of school (i.e. “primary school”). However, because children attend their closest primary school, I can determine in which year a woman gained access to a full day school based on the status of her closest primary school. Panel (a) of Figure 11 shows all the primary schools in the four states and panel (b) of Figure 11 shows the geographical distribution of full day primary schools in these states.



(a) All Primary Schools



(b) All Full Day Schools

Figure 11: Geographical Distribution of Primary Schools

This identification strategy relies on the switch-over year of a given school not being correlated with location specific factors that would affect female employment. Because I am using difference-in-difference for my estimation, this does not pose a problem, but it is still prudent to check whether the rate of reform might be correlated with location specific economic factors. In order to verify this, I look at the correlation between district-level[13] unemployment and land prices with switch-over intensity, the percentage of schools in a given district in a given year that have already converted to full day schools. Here these land prices are collected by the *Statistisches Bundesamt* and reflect the actual sale price of undeveloped land that may be developed for commercial or private use in a given year averaged at the district level. These prices are measured in Euro value of land per 100 square meters. The results of this analysis are presented in Table 25. To check the correlation, I run three simple linear regressions of a district level economic factor on the district level switch-over rate, including year dummies and district fixed effects. I cluster the standard errors at the district level.

Table 25: Economic Factors Affecting Switch-over

	(1) ols	(2) ols	(3) ols
VARIABLES	switchover_rate	switchover_rate	switchover_rate
unemployment	0.008		0.008
	(0.005)		(0.005)
landprice		0.015*	0.015**
		(0.008)	(0.007)
Constant	-0.013	-0.018	-0.031
	(0.046)	(0.012)	(0.048)
District FE	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Observations	2,076	2,206	2,036
Districts	173	173	173
R-squared	0.747	0.721	0.745

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We might be concerned that districts with low unemployment would be more likely to have a faster switch-over rate, since they require more childcare and are potentially more affluent; however, this is not observed. Column (1) shows that there is no observable correlation between the unemployment at the district level and the rate of switch-over. The coefficient on unemployment is not statistically different from zero and very small. At the same time, we might think that districts with high land prices might be economically booming and again require more childcare or have faster switch-over since they are more affluent, which is observed in Table 25. Column (2) shows that the correlation between land prices and switch-over intensity is statistically significant at the 10 percent level, but the coefficient is very close to zero. Column (3) shows that including both unemployment at the

[13] Here district refers to *Kreis*, of which there are 173 in the four states of interest.

district level and land prices does not change the correlations. As previously mentioned, these land prices are not the same as residential property prices, so they may not actually reflect the economic prosperity of a community. Because of the lack of association between unemployment and switch-over and the weak relationship between land prices and switch-over, it seems plausible that the switch-over rate of primary schools in these four states is not being driven by economic factors at the district level.

3.3 Existing Literature

The literature on childcare costs and female labour supply may be divided into structural and quasi-experimental analyses. While the structural literature in this field is vast, this work is a quasi-experimental analysis and therefore fits better in the second category; however, I highlight a key study and a thorough review paper on the structural literature relevant to this work before turning my attention to the quasi-experimental studies. Heckman (1974) provides one of the first structural models of female labour supply that considers how childcare costs impact both the household's budget constraint and the woman's indifference curves. His work is somewhat different since the childcare programmes he looks at are only eligible to women who work a certain number of hours, which is not the case in my context. Blau and Currie (2006) provide a extensive overview of models that have followed this work. They point out that the challenge with structural identification lies in exploiting differences in the price of childcare across women, something that is not always observed in the data.

The quasi-experimental studies looking at childcare and maternal labour supply are closer to this chapter because they examine the impact of a childcare policy change, which serves as a natural experiment due to the variation in price or availability across women and time periods. Lefebvre and Merrigan (2008) and Baker et al. (2008) look at the expansion of universal, highly subsidised preschool childcare in Quebec and use the rest of Canada as a control. In Quebec, childcare costs were reduced to approximately $5 per day for any children four years old and above, beginning in 1997. Lefebvre and Merrigan use the Canadian Survey of Labour and Income Dynamics (SLID) to estimate the impact this highly subsidised childcare had on mothers' labour supply. They find that the policy increased the participation of mothers who have at least one child under the age of five by eight percentage points. Their findings also show that mothers in Quebec increased their annual hours worked by 231 hours and annual weeks worked by 5.17 weeks. Both sets of results indicate changes on the extensive and intensive margins.

Baker et al. look at the same reform in Quebec and also estimate the impact of the subsidised childcare on maternal labour supply, in addition to looking at child level outcomes. They use the Canadian National Longitudinal Survey of Children and Youth (NLSCY) to look at married mothers, who still live with their husbands, due to their concern over other reforms during the same period that would have affected single mothers. Their findings show that maternal employment rates rose by 7.7 percentage points, approximately 14.5 percent of the baseline employment rate, as a result of the subsidy in Quebec. These results are very similar to the estimates produced by Lefebvre and Merrigan despite using a different data set. All of these studies differ from this work because they exploit an explicit price change across regions within a country.

Gelbach (2002) is more similar to this work since he exploits an implicit childcare subsidy as opposed to an observed change in price. He looks at the impact of public kindergarten enrolment in the United States, which varies based on birth month, on maternal labour supply. Enrolment in a public kindergarten the year before primary school begins is an implicit childcare subsidy to the family since there is no formal cost. The estimates obtained from his instrumental variable, whether or not the child qualifies for kindergarten enrolment based on birth month, show large results of the implicit subsidy. Gelbach finds that for single mothers, whose youngest child is five, their child's eligibility for free kindergarten increases their labour supply by between 6-24 percent. He notes that there is no significant impact of the policy on single mothers who have both a five year old child and a younger child, but that for married women the effect size is 6-15 percent regardless of whether or not they have an additional child under five.

Gelbach also makes the point that for any families consuming fewer or the same number of hours of childcare as the primary school day, this subsidy is a "100 percent marginal price subsidy for childcare of fixed quality" and that if the family consumes more hours of childcare than provided by the length of the school day, then "the subsidy is entirely inframarginal with respect to childcare costs" (Gelbach, 2002: 308). He points out that this will induce a kink in the budget constraint, something I also address in the Methodology section of this chapter.

Most of the literature in this field looks at pre-primary childcare, which makes my study a more unusual contribution to the field since I look at the extension of the primary school day. Contreras et al. (2010) is one of the only other papers to also look at the extension of the primary school day as an implicit childcare subsidy. They look at Chile, which also increased school hours to a full length school day beginning in 1996. This reform was undertaken on the municipal level, which allowed the authors to exploit the quasi-experimental nature of the reform across municipalities and time (Contreras et al., 2010). They acknowledge, however, that the reform in Chile was not undertaken in a random way. Municipalities that were classified as "higher risk" were also the first to receive full day schools. The authors also combine survey data, the Chilean Socio-Economic Characterisation Survey (CASEN), with administrative data to determine when women gained access to the full school day. Similar to this work, their findings showed that the reform acted as an implicit childcare subsidy, which had a positive effect on female labour supply. They find that a 1 percent increase in full day schools causes a three percent increase in the likelihood that a woman works. When they look at women by age group, they find the largest impact of the policy for the oldest group of women (50-65). They argue this is because these women tend to have older children on average, who are affected by the reform, and that younger women tend to have younger children with different childcare needs. It is still surprising, however, that the oldest group of working age women is affected the most by the reform. Their findings also show that women already employed actually decrease their working hours.

Methodologically all of these quasi-experimental papers are similar to this paper in the identification strategy they use; however, I will exploit school level variation, which reduces some of the concern surrounding the common time trend assumption and the comparability between states or regions. This differentiates my work from much of the field as many previous studies linking childcare and female labour supply compared regions within a country.

This work is also one of the first to assess the *Ganztagsschulreform* in Germany. There has been one major longitudinal study of the reform process, which collected data on a representative sample of children in full day primary and secondary schools in 14 German states (Holtappels et al. 2008). They also surveyed the parents of these pupils as part of the study. Their statistics show that 26 percent of mothers in the sample, whose children attend a full day school, reported being able to extend their working hours and 21.2 percent reported being able to re-enter the labour market (Holtappels et al. 2008). Their study does not differentiate between mothers of children in primary school and secondary school, and since there is school choice at the secondary level, it is not clear whether or not the mothers in this sample might be choosing a full day school and as a result, be able to change their labour supply. Holtappels et al. (2008) also examines the impact the reform has had on child level outcomes, something I am not able to do in this chapter due to data limitations.

Marcus et al. (2013) provide a descriptive overview of the reform using the German Socio-Economic Panel (GSOEP) to show what types of families attend full day schools, but do not identify the causal impact of the reform on female employment. They find that primary school aged children whose mothers work full time and children in single mother households are more likely to attend a full day school, and that children with immigrant parents and children from families that receive social benefits are also more likely to attend a full day school. Marcus et al. also find that over time more children from low socio-economic households are attending full day schools. They use a question in the GSOEP that asks whether or not the child attends a full day school. This is not the same as when a mother would have gained access to a full day school because the full day school might still be opt-in once a school switches over. Marcus et al. provide evidence that either some sort of selection into full day schooling is taking place once it is available at a school or that the switch-over of schools might not be random since certain groups are more likely to take up the opportunity.

Rainer et al. (2013) also use the GSOEP and propensity score matching to look at the impact of various types of childcare in Germany on female labour supply. They create a matched sample of women who report their child attends a full day school with very similar women whose children do not. They find that the full day school reform caused women who were already working to extend their hours, but did not draw women into the labour market. Rainer et al. find that most of the impact they find on hours and wages comes from women who were already working in the year before their child started attending a full day school. Their analysis includes all mothers of children aged 6-18 in the GSOEP, which again mixes primary and secondary schooling, thereby allowing for school choice. Combining these two groups does not account for selection into a full day school. Including women whose children are so much older also changes the importance of childcare within the sample. A woman whose youngest child is eight will have very different childcare needs from a woman whose youngest child is 17.

This chapter builds on the previous analysis of the German reform by exploiting a natural experiment in order to identify the causal impact of the reform on the labour supply of women with primary school aged children.

3.4 Descriptive Statistics

I combine the state-level data described in the previous section with individual level data from the GSOEP. The GSOEP is a longitudinal study of families and individuals in Germany, which includes questions on work and family and was started in 1984 (SOEP, 2013). It includes data on over 11,000 households across Germany. I use the GSOEP because it is a longitudinal study with more than sufficient information about the children in the family and allows me to access to the household's address. While it would be possible to gain information about the labour market participation status of German women from a labour force survey (e.g. the Mikrozensus) and have a larger sample size, this data would have the disadvantage of not letting me know the age of any children in the household and not providing the geocoded home addresses of respondents.

Table 26 lists the variables taken from the GSOEP as well as those I collected from states and schools for this analysis. Table 27 presents summary statistics for the variables used in the analysis as well as some additional demographic information.[14] My sample includes 6,965 women who are observed at some point during the period 2000-2012 in the four states of interest. I look at these years because this is the period in which the reform is taking place in the states for which I have data. I only include women who are aged 15-64, as this is the working age population as defined by the German Federal Employment Agency.

Because of attrition and sample refreshment, the GSOEP is not a balanced panel, which should be kept in mind when analysing the descriptive statistics presented in Table 27. Since I only look at

[14]Some of the variables I do not actually include in the analysis, due to the inclusion of individual fixed effects, but they prove interesting for descriptive purposes.

Table 26: Variable Descriptions

Variable Name	Description	Source
employed	Binary variable for whether or not the woman is working	GSOEP
fulltime	Binary variable for whether or not the woman is working full time	GSOEP
parttime	Binary variable for whether or not the woman is working part time	GSOEP
weeklyhrs	Continuous variable for the weekly hours worked	GSOEP
lnweeklyhrs	Continuous variable for the logarithm of weekly hours worked	GSOEP
treatment	Binary variable for whether or not the woman has primary school aged children and access to a full day school	GSOEP and state-level
accessgts	Binary variable for whether or not the woman has access to a full day school	GSOEP and state-level
age	Continuous variable for woman's age	GSOEP
pschildren	Binary variable for whether or not there are primary school aged children present in the household	GSOEP
preschoolkids	Binary variable for whether or not there are preschool aged children present in the household	GSOEP
olderkids	Binary variable for whether or not there are children aged 11-15 present in the household	GSOEP
labinc	Continuous variable for the gross labour income of the individual	GSOEP
lnlabinc	Continuous variable for the log gross labour income of the individual	GSOEP
yrsedu	Continuous variable for woman's years of education	
married	Binary variable for woman's maritial status	GSOEP
singmom	Binary variable for whether or not woman lives alone with children	GSOEP
husbandwork	Binary variable for whether or not male partner is employed	GSOEP
syear	Year dummy variables	GSOEP
state	State dummy variables	GSOEP
pid	Unique person identifier	GSOEP
schoolid	Unique school identifier	state-level

Table 27: Summary Statistics By Observations

Variable Name	**N**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
employed	36,158	0.602	0.489	0	1
fulltime	22,710	0.408	0.492	0	1
parttime	22,710	0.592	0.492	0	1
weeklyhrs	20,993	30.993	13.495	1	80
lnweeklyhrs	20,993	3.296	0.596	0	4.382
labinc	21,663	1,776	1,652	0	99,999
lnlabinc	21,542	7.160	0.874	0	11.513
treatment	36,158	0.012	0.146	0	1
accessgts	36,158	0.110	0.314	0	1
age	36,158	40.2	13.8	15	64
pschildren	36,158	0.133	0.339	0	1
preschoolkids	36,158	0.104	0.305	0	1
olderkids	36,158	0.207	0.405	0	1
yrsedu	31,141	12.065	2.627	7	18
married	36,158	0.554	0.497	0	1
singmom	36,158	0.165	0.371	0	1
husbandwork	26,713	0.796	0.403	0	1
syear	36,158			2000	2012
state	36,158			1	4

NB: N are person-year observations

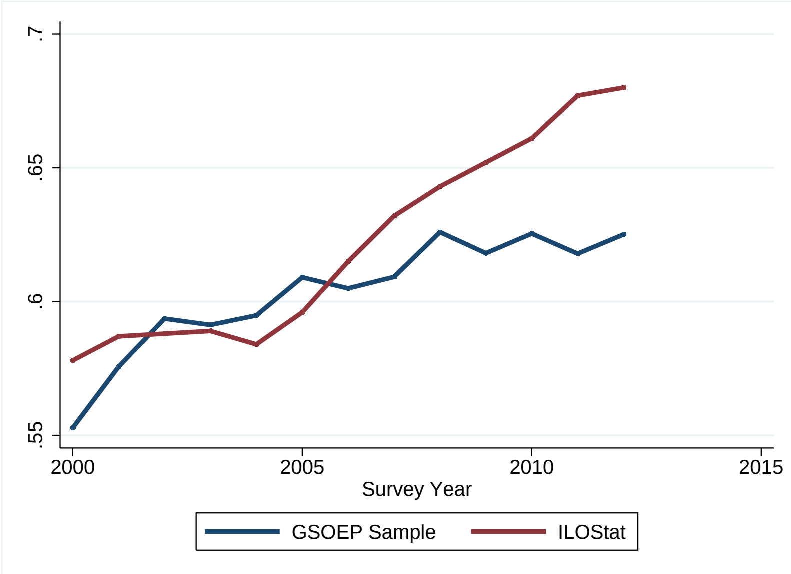


Figure 12: Employment Ratio Within Sample and in ILOStat Data By Year

women up to age 64, once a woman turns 65, she also drops out of my sample. This means that I have at most 36,158 person-year observations for some variables. There are fewer observations for some variables due to missing values.

The GSOEP is a representative sample of people living in Germany and the descriptive statistics in my sample confirm this. Over half of the women I observe in this sample are married and on average 40 years old. Twenty percent of the sample have children in secondary school, thirteen percent have primary school aged children, and only ten percent have pre-school aged children, potentially due to declining fertility across time. On average, the women in my sample have completed approximately twelve years of education, which is more than the legally required minimum of nine years of education. A university educated woman will have completed 18 years of education, the maximum reported in the data.

Across all time and person observations, 60 percent are employed. The dynamics of their employment proves similar to the ILO statistics presented earlier; over the period of interest, employment has also increased for the women in my sample (Figure 12). This indicates a change at the extensive margin of the labour market over the period of interest not only in the ILO data, but also in my sample.

For the women that work, their average hours, "weeklyhrs" in Table 27, in the sample is approximately 31. This is below the threshold of 35 hours[15] for a full time job, indicating that many women in the sample are engaged in part time work, which is confirmed by the distribution of hours worked presented earlier in the section on Female Labour Supply in Germany (see Figure 9). The variables "parttime" and "fulltime" confirm this. Nearly 60 percent of the women working

[15] Here I use the OECD definition of part time work to be anything less than 35 hours per week (OECD Glossary of Statistical Terms).

Table 28: Employment Information by Number of Primary School Children

	Employed		
	Mean	N	Standard Deviation
Number of Primary School Children			
0 children	0.625	27,721	0.484
1 child	0.574	5,158	0.495
2 children	0.472	2,701	0.499
3 children	0.367	528	0.483
4 children	0.319	47	0.471
5 children	0.333	3	0.577
	Weekly Hours of Work		
	Mean	N	Standard Deviation
Number of Primary School Children			
0 children	32.903	16,720	13.043
1 child	24.899	2,857	12.688
2 children	21.117	1,219	12.001
3 children	18.616	183	11.876
4 children	16.538	13	7.957
5 children	10	1	

NB: N are person-year observations

are working in a part time capacity.

For the women that work, their monthly labour income is reported in the variable "labinc," with its natural log reported in "lnlabinc." Income in the GSOEP is top coded with 99,999, which affects the mean value reported in Table 27. The median income for the women working in my sample is 1,500 Euros per month, which is less than the reported mean.

If we believe that mothers have different labour force participation patterns as a result of having children, then we would expect to observe a difference in the employment and hours data for mothers and non-mothers. As Table 28 shows, there is a difference in the employment rate of women who have primary school aged children versus those who do not. Furthermore, the employment rate decreases as the number of primary aged school children in the house increases. Table 28

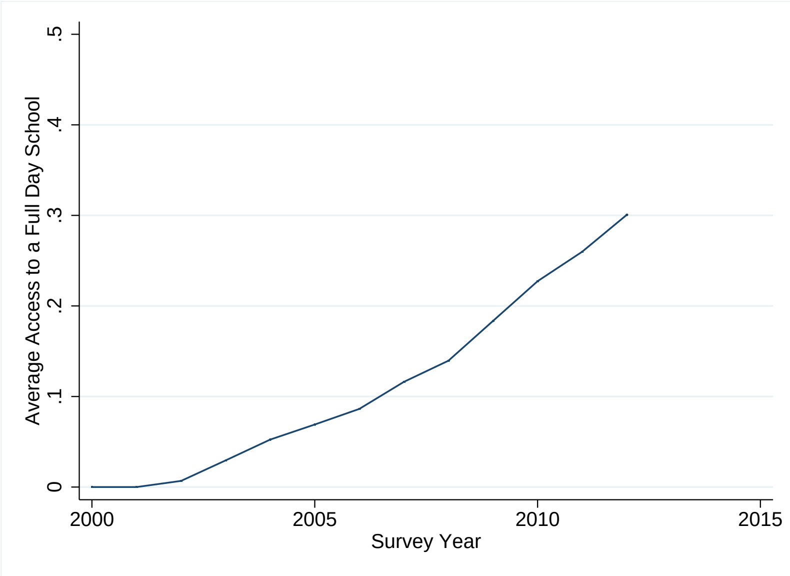


Figure 13: Proportion of Sample with Access to Full Day Schools By Year

shows a similar trend for weekly hours of work: they decrease as the number of primary school aged children in the household increases.

As previously mentioned, during this period, 13.3 percent of my person-year observations have primary school aged children. This limits the number of women in the sample affected by the policy significantly since having access to a full day school only matters if you have primary school aged children. Because of this, my "treatment" is the interaction of the binary variable for having access to a full day school ("access") and the binary variable for having primary school aged children ("pschildren"). This means that for a woman to be treated, or affected by the reform, she must not only have access to a full day facility, but also have primary school aged children. When interpreting my results, the coefficient of interest on the variable treatment, will tell me the additional impact of gaining access to a full day school when a woman has primary school aged children.

Figure 13 shows what percent of the sample of 6,965 women has access to a full day facility in any given year and Table 29 shows the exact number of women in the sample in a given year and state that have access and of those women, which ones are treated, i.e. also have primary school aged children.

Table 29: Number of Individuals with Access and Treatment By Survey Year

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	Overall
Number of women with access	0	0	21	85	144	183	245	310	344	501	551	746	865	1,413
...as a proportion of sample	0	0	0.007	0.030	0.052	0.069	0.086	0.116	0.141	0.186	0.228	0.260	0.299	0.203
...in Bavaria	0	0	0	0	0	0	15	34	34	136	151	191	281	407
...in Hesse	0	0	0	39	47	50	74	84	109	131	183	223	248	405
...in Rhineland-Palatinate	0	0	21	45	61	75	94	100	103	119	108	157	164	287
...in Schleswig-Holstein	0	0	0	1	36	58	62	92	98	115	109	175	172	322
Number of treated women	0	0	1	12	21	20	29	39	35	49	61	86	89	211
...as a proportion of sample	0	0	0	0.008	0.012	0.015	0.017	0.025	0.031	0.038	0.042	0.048	0.054	0.045
...in Bavaria	0	0	0	0	0	0	3	8	11	14	17	21	26	55
...in Hesse	0	0	0	5	8	12	13	13	11	9	20	25	25	62
...in Rhineland-Palatinate	0	0	1	6	9	4	9	11	8	12	8	14	14	43
...in Schleswig-Holstein	0	0	0	1	4	4	4	7	5	14	16	26	24	52
Number of women in panel	3,124	2,769	3,103	2,887	2,747	2,650	2,840	2,669	2,465	2,731	2,424	2,871	2,878	6,965

This table shows that access and treatment are both increasing over time as more schools switch-over to become full day facilities, but that treatment has been limited. I only observe a total of 211 women who have gained access to a full day school whilst they had primary school aged children. This is the result of only having school data on four states, the slow switch-over rate of schools within those states, and the limited number of women that have primary school aged children at the same time they gain access. Bavaria, the most populous state in my data set, has had the slowest switch-over rate, limiting the absolute number of treated women. Nevertheless, 211 women will still allow me to estimate the impact of the extension of the school day on female labour supply.

3.5 Methodology

3.5.1 Theoretical Models of Labour Supply and Piecewise Linear Budget Constraints

Models of labour supply include an agent maximising her utility function subject to a budget constraint. As Cahuc and Zylberberg (2004) point out, agents face "at best" a piecewise linear budget constraint when considering their labour supply. This piecewise linearity arises from complexities in the fiscal system as well as costs to entering the labour market. For the purposes of this paper, I consider two simple, static labour supply models, both with piecewise linear budget constraints. The first model has a fixed cost to entering the labour market, which is the cost of childcare beyond the hours provided by primary schooling. The second model looks at a scenario where mothers pay childcare costs as a variable cost, deducted from their hourly wage, and it depends on the number of hours they work. This gives rise to a kinked budget constraint.

Both models also include a small fixed cost to working at all, which we may think of as a commuting cost. This is a necessary cost to impose, and standard in most models, because without, women who are not working might never be drawn into the labour market as a result of the school day extension. The size of this fixed cost could vary, however, for the purposes of these graphs, I have left it as relatively small and the same across all graphs.

In practice, women probably face both fixed and variable childcare costs to working, depending on the types of childcare arrangements they choose, yet I consider them separately for the sake of simplicity. Regardless of which model is chosen, however, they make very similar predictions and both unambiguously predict that women will never leave the labour market as a result of this policy change.

I present all the results of the two models graphically although of course they could also be solved mathematically. The graphical presentation, however, proves more informative and allows me to circumvent the complexities that would arise from optimisation with a piece-wise linear budget constraint.

Fixed cost to working The first model shows the effect of fixed costs on labour supply and is similar to the static model in Pencavel (1987): the women in this model want to maximise their utility function $U(x,\ h)$, where x represents consumption and h represents hours of work. This utility function is well-behaved (real valued, continuous, and quasi-concave).

As the women face a fixed cost to entering the labour market, childcare, the budget constraint will be piecewise linear. This is because I assume that once a woman decides to work any number of hours beyond $\hat{h}$, she must pay for childcare. We may think of this fixed cost as the money paid to a childminder or nanny, or the set amount of money that a childcare programme charges per week or month. The budget constraint for working therefore becomes:

$$px = \begin{cases} wh + \tilde{y}, & if\, h \leq \hat{h} \\ \\ wh + \tilde{y} - F, & if\, h > \hat{h} \end{cases} \tag{11}$$

Where p is the price of consumption, which may be normalised to 1, w is the wage the woman receives for working, which I assume to be the same across hours worked, $h > 0$ are the hours she works, $\tilde{y}$ is her non-labour income minus the general fixed cost to working ($\tilde{y} = y - z$), and F is the fixed cost of childcare needed outside of the primary school day. Combining these two budget constraints gives us the piecewise linear budget constraint in Figures 14. If the woman does not work at all, her income will be composed of non-labour income, y, only. If she works any hours less than $\hat{h}$, she has to pay a small fixed cost of working, z, e.g. transport costs, which explains why the first linear segment of the budget constraint begins below the full amount of non-labour income y. If she works any hours above $\hat{h}$, however, not only does she have to pay the fixed cost of working at all, but she must also pay the fixed cost of childcare beyond the hours in which the child is at primary school.

The woman solves the following maximisation problem for each segment of the budget constraint to determine her participation and her hours:

$$max_{x,h} \quad U(x,\ h) \quad s.t.\ px = \begin{cases} wh + \tilde{y}, & if\, h \leq \hat{h} \\ \\ wh + \tilde{y} - F, & if\, h > \hat{h} \end{cases} \tag{12}$$

Solving this maximisation problem yields tangency conditions, which is how the woman will determine her participation and hours. There are three different cases to consider: women who do not work, women who work $\hat{h}$ hours, and women who work more than $\hat{h}$. I do not consider the case of women who work less than $\hat{h}$ hours or who work exactly $\hat{h}$ hours at an interior solution because provided they have well-behaved preferences, they will not be affected by the reform.

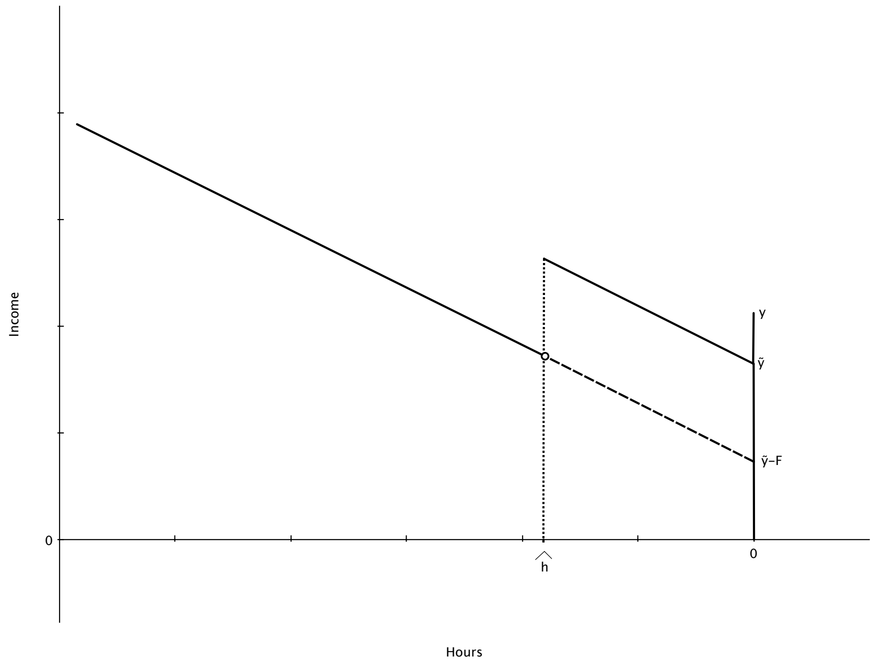


Figure 14: Piecewise Linear Budget Constraint with Fixed Cost of Childcare

The first case is that of a woman who was not working before the extension of the primary school day. This scenario may be seen in Figure 15. The woman is initially working zero hours at point A, which is a corner solution, because this same indifference curve is not tangent to any attainable point on her budget constraint. By not working, this woman is on her highest attainable indifference curve. Once the school day is extended and $\hat{h}$ increases to $\hat{h}'$, the woman is able to move onto a higher indifference curve at the newly attainable portion of her budget constraint, which changes her employment status from not working to working. This is shown by the new tangency point B. This result relies on the general fixed cost to working, e.g. the commuting cost, because without it, no women would be drawn into the labour market as a result of the decrease in childcare costs.

The second case is a woman who was working exactly $\hat{h}$ hours before the reform. In this case, the woman has coordinated her work schedule exactly with that of the primary school. Figure 16 shows the budget constraint and indifference curves for the woman who is at a corner solution at $\hat{h}$. Initially she is working $\hat{h}$ hours, at point A. The discontinuity in the budget constraint coupled with her preferences brings about this corner solution. Once school hours are extended to from $\hat{h}$ to $\hat{h}'$, this woman is able to move onto a higher indifference curve on the newly extended portion of her budget constraint, at the tangency point B. This woman is now working more hours than before.

The third case, shown in Figure 17 is a woman who is working more than $\hat{h}$ hours. This would include all mothers who work full time and some mothers who work part time, provided their hours go beyond the length of the primary school day. Before the reform, this woman determined her participation and hours by selecting point A, where her highest indifference curve is tangent

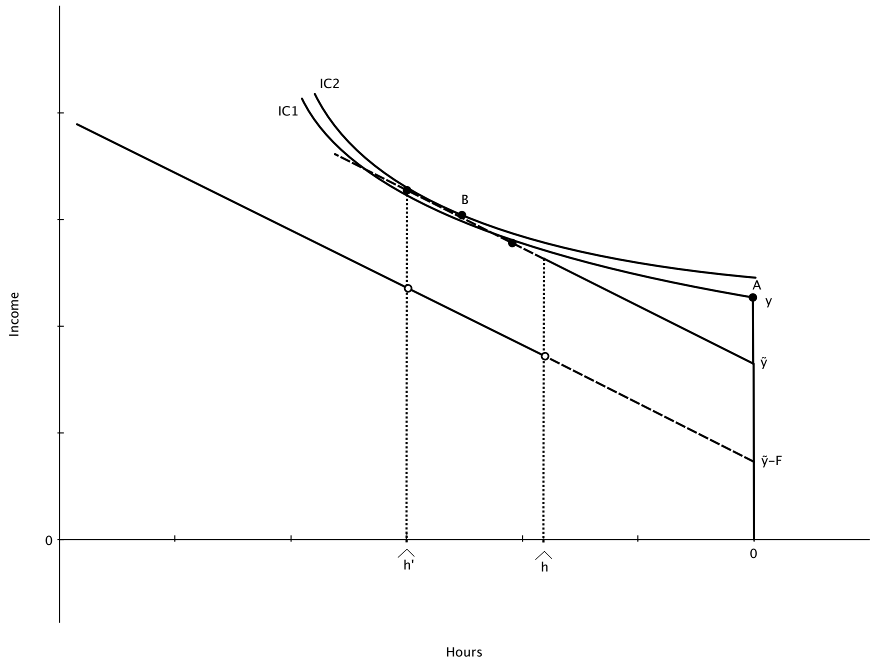


Figure 15: Case 1: Woman not working

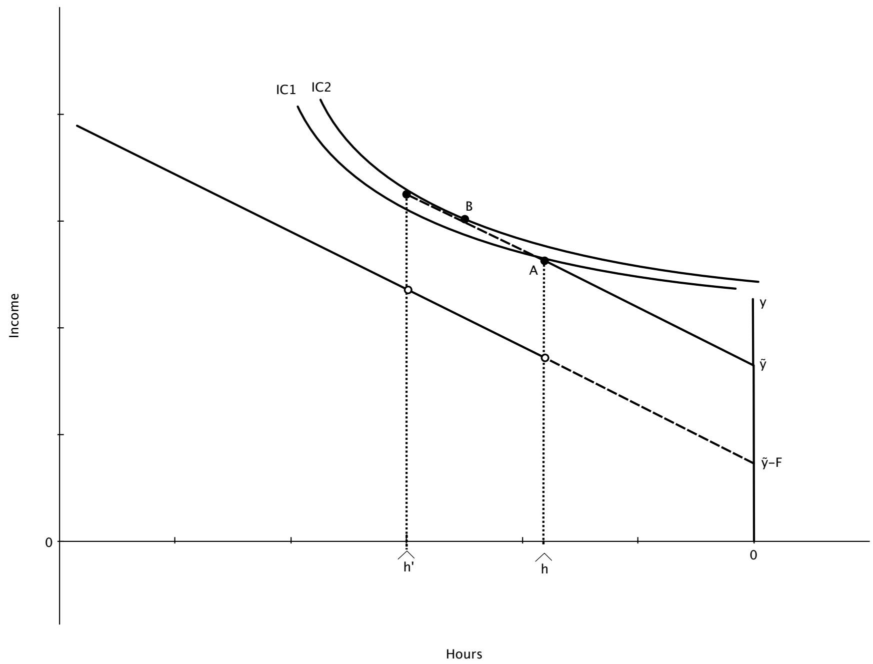


Figure 16: Case 2: Woman working exactly primary school hours

to her budget constraint. Once school hours are extended to from $\hat{h}$ to $\hat{h}'$, this woman has an extended portion of the budget constraint at which the fixed cost of childcare is not deducted. As a result, she can move onto a higher indifference curve, which is tangent to her newly extended budget constraint at point B. In this case she has decreased her working hours, but is still better off than before.

The outcome in this third case depends on several factors. First, the size of the fixed cost to

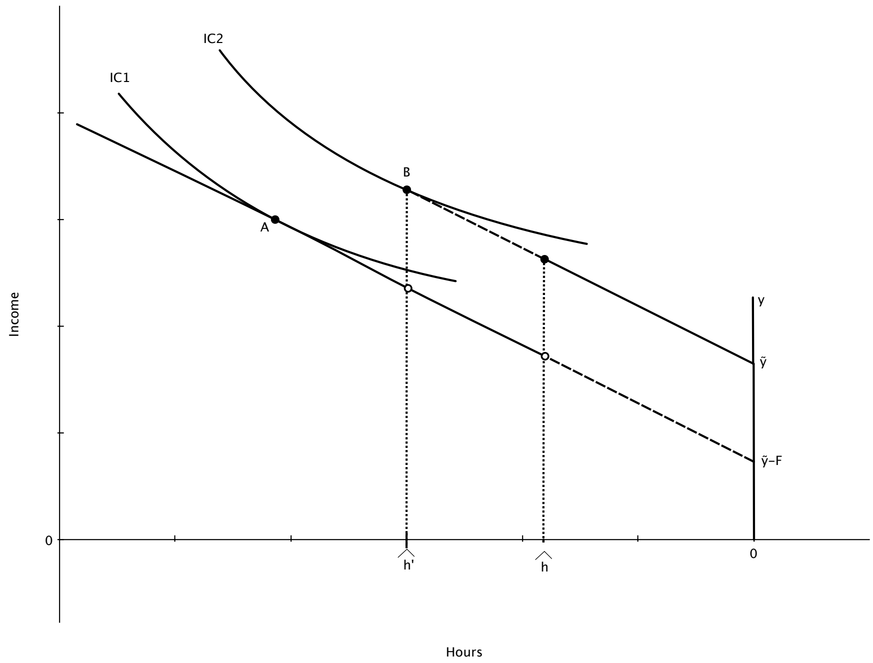


Figure 17: Case 3: Woman working beyond primary school hours

working. If the fixed cost to working is very small, which is probably not the case with childcare, and the wage rate is high, meaning the budget constraint is steep, then the woman may not experience an income effect as a result of the reform. This would apply to women working a lot of hours in this case. The woman who does not experience an income effect as a result of the reform has no reason to change her working hours. Second, whether or not leisure is a normal good. If leisure is a normal good, then the woman facing this labour supply problem must stay at the same hours or decrease her hours as a result of the reform because she only faces an income effect in this case.

This model predicts that as a result of the extension of the primary school day, mothers who were not working before will be drawn into the labour market, but that for women who were already working before the reform, they will either increase or decrease their working hours. The sign of this change depends on how many hours they were working before. If we observe mothers setting their working hours exactly to match the length of the primary school day, then we would expect them to increase their hours as a result of the reform. For women working more hours than the length of the school day, this model predicts that they will shorten their working hours. In all cases, women are able to reach a higher indifference curve as a result of the policy change, making them better off than they were before.

Variable cost to working The second static model includes childcare as a variable cost to working, which will be deducted from the hourly wage. Instead of being deducted as a fixed cost, we can think of mothers paying for childcare only for every hour they work beyond the length of the primary school day. The variable cost of childcare could be the hourly rate that a mother must pay to a childcare facility to watch her children after the school day ends and depends on how much she works. The budget constraint for mothers in this model is:

$$px = wh + \tilde{y} \tag{13}$$

Where

$$w = \begin{cases} w_1 & if\, h \leq \hat{h} \\ w_2 & if\, h > \hat{h} \end{cases} \tag{14}$$

$$w_1 > w_2 \tag{15}$$

This budget constraint is drawn in Figure 18.

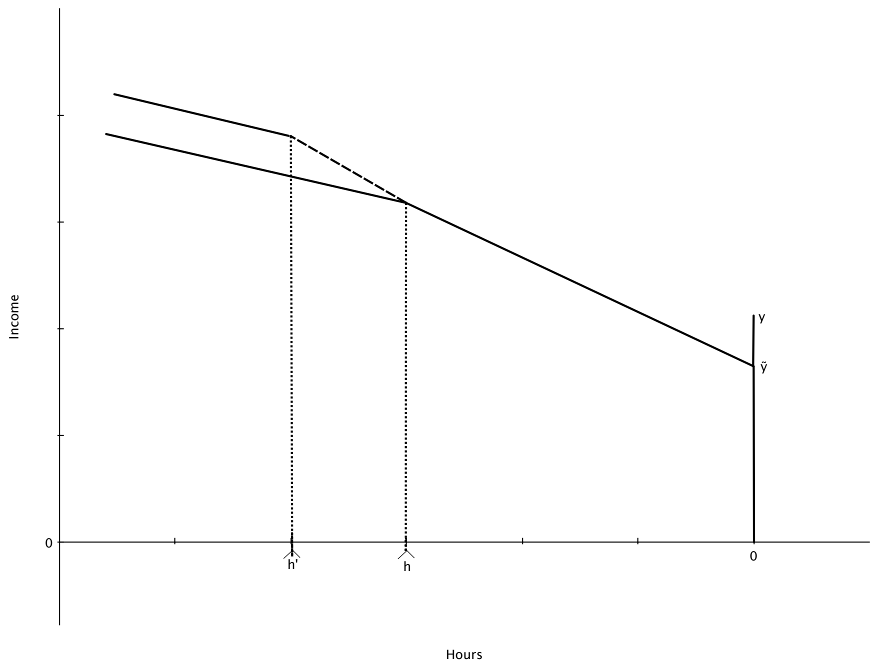


Figure 18: Piecewise Linear Budget Constraint with Variable Cost of Childcare

Here the cost of childcare beyond $\hat{h}$ hours, the hours the child spends in school, are deducted from the woman's wage, which is why w_2 is less than w_1. This will create a kink in the budget constraint as the relative price ratios change. Again, I normalise $p = 1$ for the sake of convenience. As in the fixed cost case, there is a fixed cost to working, z, such that non-labour income is y when the woman does not work at all and $\tilde{y} = y - z$, when she works more than zero hours. Again, this fixed cost to working could be a commuting cost.

As in the fixed cost case, this labour supply problem may be solved by maximising the utility function given the budget constraint and the solution will provide us with a similar tangency condition. The problem similarly looks like this:

$$max_{x,h} \quad U(x,\, h) \quad s.t. \;\; px = \begin{cases} w_1 h + \tilde{y}, & if\, h \leq \hat{h} \\ \\ w_2 h + \tilde{y}, & if\, h > \hat{h} \end{cases} \tag{16}$$

As in the fixed cost model, I consider the same three cases: women who do not work, women who work $\hat{h}$ hours, and women who work more than $\hat{h}$ hours. Similarly, women who work less than $\hat{h}$ hours will not be affected by the reform, provided they have well-behaved preferences, so I do not explore this case.

The first case is that of a woman who is not working before the school day gets extended. This may be seen in Figure 19. Here the woman was choosing not to work at a corner solution, point A, because her highest possible indifference curve was not tangent with an attainable portion of her budget constraint. Once school hours were increased, however, she is able to choose point B on a higher indifference curve, thereby entering the labour market and increasing her utility. As in the childcare as a fixed cost model, the general fixed cost to working underpins this result. Without this additional fixed cost, the decrease in childcare costs would not induce these non-working mothers to enter the labour market.

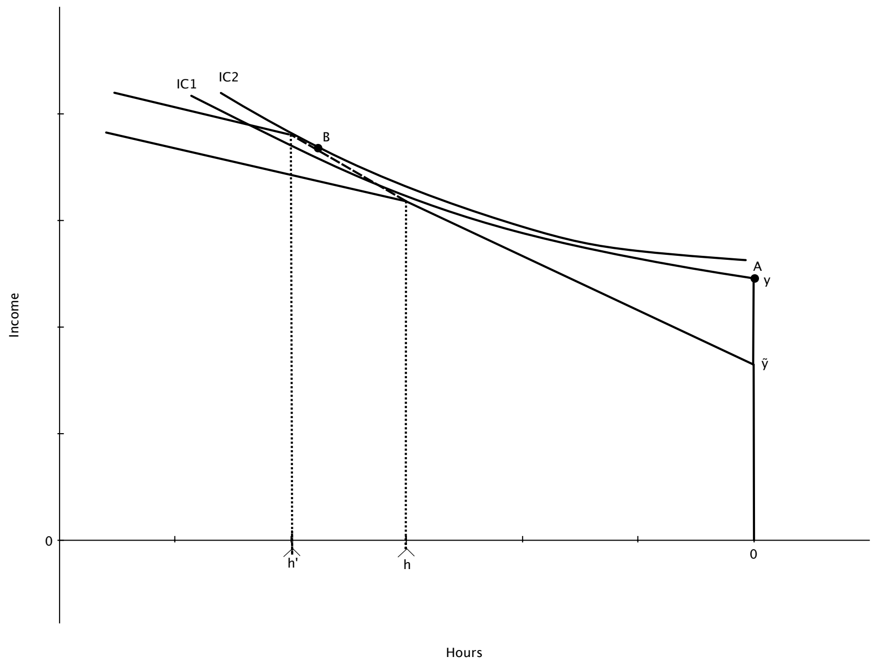


Figure 19: Case 1: Woman not working

The second case of a woman working exactly $\hat{h}$ hours mirrors the fixed cost model in that there are two possible outcomes, which depend on her indifference curves. The first outcome is the less interesting outcome. Here the woman's indifference curve is tangent to her budget constraint at the kink point, but the shape of the indifference curve is such that when school hours are extended and the budget set gets larger, her initial choice is still her optimal choice. In this case the woman does not extend hours and her utility does not change. This is the same scenario for a woman who is initially working less than $\hat{h}$ hours, given that her preferences are well-behaved.

Figure 20 shows what happens when her indifference curve has a different shape and the policy change causes her to change her working hours. Initially this woman is also at point A, at the kink point, however, once school hours increase and her budget set gets larger, she can move onto a higher indifference curve. This new tangency point, at point B, is possible because of the increased budget set. This woman is now working more hours and has higher utility.

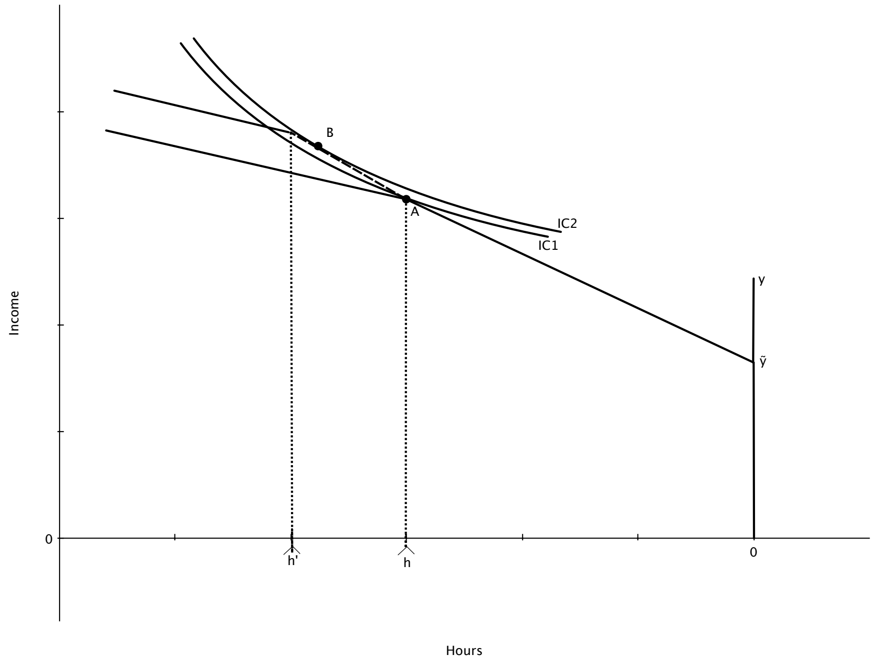


Figure 20: Case 2: Woman working exactly primary school hours

The third and final case is that of a woman working hours above $\hat{h}$. This situation is depicted in Figure 21. This woman was working some number of hours beyond the length of the primary school day. When the school day was extended, her budget set also expanded due to the income effect. This resulted in her being able to choose a point on a higher indifference curve, point B in Figure 21. As a result of moving from point A to point B, this woman has decreased her hours, but increased her utility. This decrease in hours is a direct result of the income effect being the only effect coupled with labour being a normal good.

The same decrease in hours and increase in utility may be shown for values of h greater than, but closer to $\hat{h}$ in magnitude, as seen in Figure 22. In this case we observe an income and a substitution effect, which work in opposite directions. This is the only case where the woman will experience a substitution effect as the wage rate has changed. In Figure 22, the income effect dominates the substitution effect and the woman decreases her hours worked. The opposite would also be possible if the substitution effect dominated.

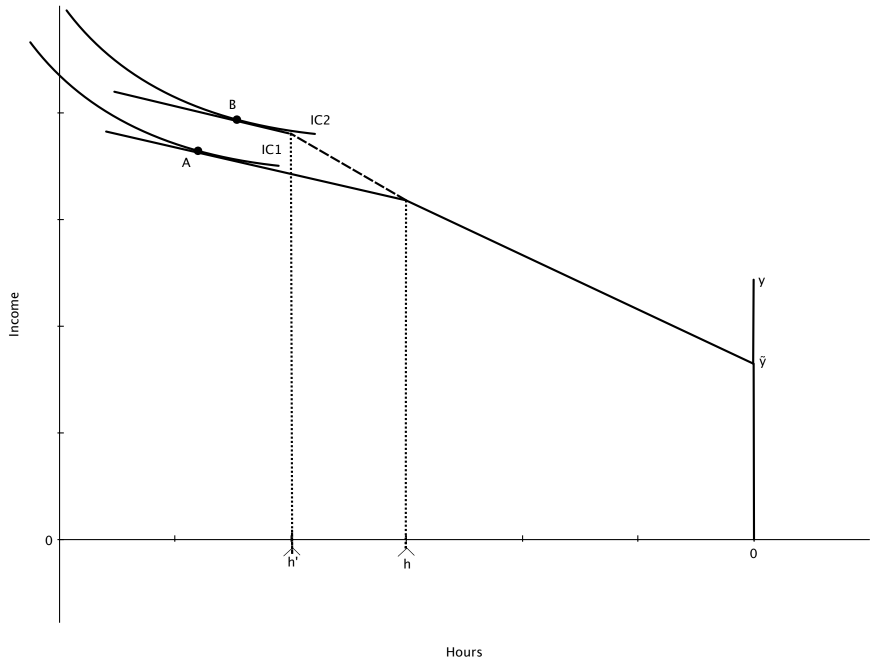


Figure 21: Case 3a: Woman working beyond primary school hours

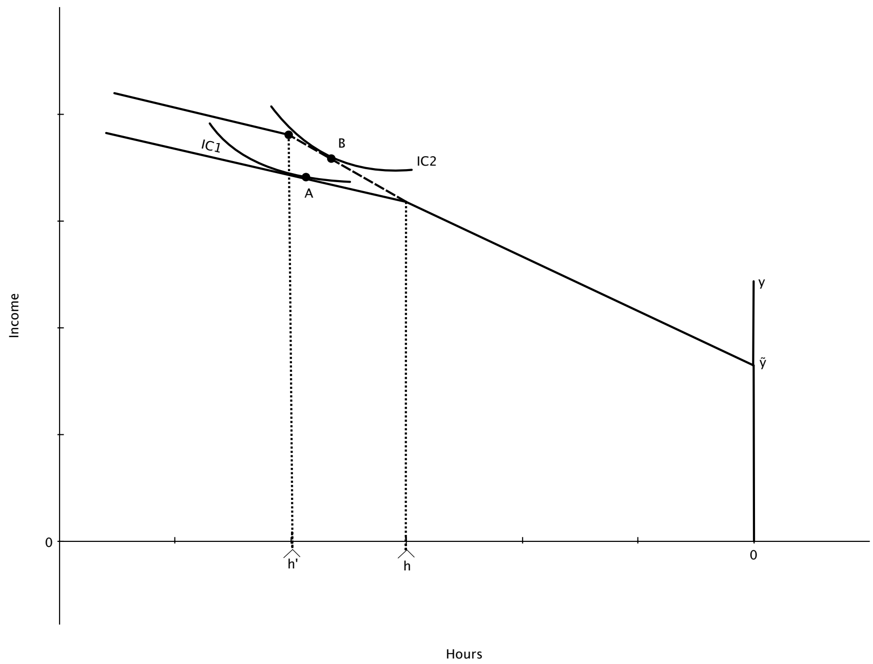


Figure 22: Case 3b: Woman working beyond primary school hours

3.5.2 Identification

Overall the variable cost model predicts the same outcomes as the fixed cost model. The extension of the primary school day will cause women not working to enter the labour market, change at the extensive margin, but the effect of the reform on women already working is unclear. They will either increase, decrease, or remain at the same number of hours. This intensive margin change is unclear due to the different responses of individuals depending on the number of hours they work.

In almost all cases, however, the policy change allows women to reach a new utility maximising choice on a higher indifference curve.

Issues In this chapter, I attempt to identify the impact of the extended school day on female labour supply. Usually, the challenge to identification in this type of research lies in disentangling the endogenous work and childcare decisions. The advantage of the German reform is that access to a full day school comes exogenously to different women at different times. I exploit this variation in my identification strategy (which is described in more detail in the section Data on the Reform).

The exogeneity of the reform means that women cannot affect when they gain access to a full day school because their children must attend the closest primary school. There is still the possibility that some families may send their children to a private school that offers extended hours or potentially move house to live closer to a full day primary school. Unfortunately I am not able to identify reasons for moving house, however, it does not appear to pose a serious problem to identification. I observe 137 women in the data set who have moved house from a home where the closest primary school was not a full day school to a home where the closest primary school is a full day school; however, only 7 of these women moved whilst having primary school aged children, thereby changing their treatment status. This is a small percentage of my treated sample, but something I will address in the Robustness section.

Verification The GSOEP includes a limited number of variables related to child level outcomes, which may have been affected by this reform. It does not include enough information on grades or other academic outcomes to assess the impact of the reform on learning outcomes; however, there are some limited time use variables, which will allow me to verify the validity of my identification strategy. My goal is to show that the children of the woman I am assigning treatment to have actually experienced a change due to the treatment and therefore, assigning treatment to their mothers is a valid approach.

The GSOEP collects information on how many hours primary school aged children spent in various types of childcare, including hours spent at school, when the children start primary school (age six) and again shortly before they transition to secondary school (at age ten). These questions are answered by parents, and the data is only available for a rather small sample of children, 152, who live in the four states of interest.[16]

If the identification strategy is working, we would expect that having access to a full day primary school would increase the number of hours a child spends at primary school. I use the same variable,

[16]It should be noted, however, that these results are being estimated on a very small sub-sample of the 152 children, as many of them have missing values on the outcome variable in one year of being surveyed and therefore drop out of the fixed effects estimation. This may be seen in Table 31.

Table 31: Estimates on Child's School Hours

VARIABLES	hrsschool
FDS	16.268***
	(2.095)
Constant	14.665***
	(0.532)
Individual FE	Yes
Year dummies	Yes
Observations	181
R-squared	0.189
Number of children	152

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

access to a full day school, as determined by proximity to closest primary school, to estimate the following model:

$$SHours_{it} \quad = \quad \alpha_0 + \beta \, FDS_{it} + \gamma_i + \theta_t + e_{it} \tag{17}$$

Here the subscript "i" denotes the child and the subscript "t" denotes the year. The variable $SHours_{it}$ is the number of hours the child spends in school and FDS_{it} is the binary indicator for whether the child's closest primary school is a full day school. Indeed, as the results in Table 31 show, having access to a full day school increased the number of hours a child spent per week at primary school by approximately 16 hours. Given the confidence interval on this coefficient, this result is in line with an extension of the school day by 2.5 hours per day.

Based on this analysis, it seems as though the strategy of using the closest primary school to determine access to a full day facility is a valid method for determining treatment status of mothers.

3.5.3 Empirical Strategy

In this section, I describe the models used to estimate the effect of the policy on maternal labour supply. I estimate two main models: one looking at changes in employment status and one looking at changes in hours. This allows me to explore the impact of the policy on the extensive and intensive margins. My variable of interest is "treatment," which is the interaction of whether or not a woman in the sample has access to a full day school and primary school children. For this reason, I always include the binary variables for access and primary school children in all regressions.

The model for employment status, whether or not the woman is employed, takes the following form, where E_{it} is a binary variable that takes the value "1" when the woman is employed and "0" otherwise:

$$E_{it} \quad = \quad \alpha_0 + D_{it}\delta + \eta_i + \phi_{st} + X_{it}\beta + \varepsilon_{it} \tag{18}$$

All of the employment models are estimated either as linear probability or conditional logit models because of the binary outcome measure. This model does not take into account whether or not the woman is working part or full time, but rather pure, binary employment status. In all models, "i" signifies "individual" and "t" signifies "year." This specification allows for the inclusion of the treatment variable, D_{it}, which can switch back and forth between 0 and 1 depending on the woman's treatment status. These regressions also include an individual fixed effect, η_i, state-year dummies, ϕ_{st}, as well as standard errors, ε_{it}, clustered at the individual level. The individual fixed effects pick up any individual specific, time invariant characteristics that could explain employment status. Similarly, the state-year dummies should explain any variance in employment status caused by events occurring in a specific year in the state of residence, i.e. larger macroeconomic events. Since individual decisions to supply labour could be correlated over time, it makes sense to cluster at the individual level even though the treatment is occurring at the school level.

The variables included in the vector X_{it} include a binary variable for whether or not the woman has primary school aged children, a binary variable for whether or not she has access to a full day school, a binary variable for whether or not the woman has preschool aged children, and a binary variable for whether or not she has children in secondary school. These variables are included in order to disentangle the general effects of being a mother on labour supply that are not affected by this policy. These covariates are described in Table 26 and summarised in Table 27.

Similarly, the regressions exploring weekly hours of work, take the following general form:

$$H_{it} \quad = \quad \alpha_0 + D_{it}\delta + \eta_i + \phi_{st} + X_{it}\beta + \epsilon_{it} \tag{19}$$

Here H_{it} is a continuous variable representing either level hours of work or the logarithm of hours worked. The vector X_{it} includes the same covariates as in the employment regressions and again, the standard errors, ϵ_{it}, are clustered at the individual level.

I estimate the employment and hours models separately as opposed to in a joint participation-hours framework because I am not working in the standard censored context. In my data set, all of the hours worked are positive values; any woman who does not work receives a missing value instead of a zero for her hours. This allows me to estimate the impact of the policy on hours conditional on employment, which is the intensive margin. I still look at how the extension of the school day affects the extensive margin by looking at the dummy variable for employment status. By separating the two, however, I am able to disentangle the extensive from the intensive margin for more nuanced analysis.

This empirical strategy, however, does not take potential spillover effects into account. There are a limited number of jobs available in the labour market and in order for these mothers to enter the labour market, vacancies must be created at a fast enough rate or some other workers must be getting squeezed out or having their hours reduced. The workers who exit the labour market could be women who do not have primary school aged children or men. Since this reform was widely discussed in Germany, it is likely that endogenous job creation took place as firms created new jobs in response to the reform. These are all things to keep in mind when thinking about the policy implications of this type of reform.

3.6 Results

All regressions presented in this section include the full sample of all women aged 15-64, who live in the four states of interest during the period 2000-2012. Because the treatment variable is defined as the interaction of the variable for access and the variable for having primary school aged children, the control group is actually composed of two different sub-groups: women who have primary school aged children, but do not have access to a full day school, and women who do not have primary school aged children, but do have access to a full day school. In order to disentangle the control group, I also run the same models on a sub-sample of only women with primary school aged children. These results follow a discussion of the main employment and hours regressions on the entire sample.

3.6.1 Full Sample

In Table 32 I present the results from the regressions on employment using the full sample. I first run a linear probability model on the binary outcome variable in Column (1), followed by a conditional logit in Column (2). Column (1) shows that being treated, having access to a full day school when you have primary school aged children, increased the probability of being employed by 6.9 percent.[17] This effect is statistically significant at the five percent significance level. The marginal effect associated with the logit coefficient on the treatment variable reported in Column (2) is 0.075, which is comparable to the estimate obtained through OLS.

Using a back of the envelope calculation, I estimate that this 6.9 percent increase in the probability of being employed translates into a less than one percentage point increase in overall female employment over this period. This is rather small in terms of a change at the extensive margin in the macro picture since over the period 2000-2012, female employment in Germany increased by approximately 10 percent (ILOStat).

[17] This marginal effect on labour force participation is (0.071-0.002), the difference between the coefficient on the treatment variable and the coefficient on the access variable. The coefficient on the treatment variable, 0.071, measures the reduction in the negative impact of having primary school aged children on labour market participation from gaining access to a full day school.

The potential of this 6.9 percent effect, however, should be considered. As my data on primary schools shows, only 50 percent of primary schools in these four states have switched over to full day schools. If this treatment effect remains constant, scaling up the size of the treatment by switching over all primary schools could increase overall female employment by seven percentage points, which would be very significant in the German context.

The sign and significance of this coefficient do not change when I move from the linear probability model to the logit regression. Here we observe a small, yet statistically significant effect of providing women with an implicit childcare subsidy on their labour supply. Both of these regressions include individual fixed effects and state-year dummies with standard errors clustered at the individual level.

As might be expected, having primary school aged children, pre-school aged children, or secondary school aged children all decrease the probability of being employed. There is clearly a negative effect of being a mother on employment, which matches the previously discussed statistics of West German mothers' employment patterns. Once women have children they exit the labour market and often do not return.

All of these results are being estimated on the "switchers," women who changed their labour market status as a result of the treatment. This is because of the nature of the difference-in-difference estimation; we are interested in how the treated individuals' employment status changes across periods compared to the non-treated sample's changes. In the case of the conditional logit model, women whose employment status does not change across periods are actually dropped in the estimation. This is why both the number of observations and the number of individuals are lower than the numbers reported in the Descriptive Statistics in Table 27. This reinforces the point that women are being drawn into the labour market and there is actually a change at the extensive margin as a result of this reform. This change at the extensive margin was predicted by both the both the fixed and variable cost models presented in the Methodology section.

Turning to the intensive margin, I find no impact of the reform on hours worked. These regressions are only being estimated on women who report a positive number of hours worked with zero hours being treated as missing. This means that any changes in hours will reflect changes at the intensive margin. As the coefficients on the treatment variable in Table 33 show, the effect on both level and log hours is small and not statistically different from zero. For women who were already working before getting treated, their treatment did not cause them to change their working hours. This could be driven by rigidities in the labour market that do not allow workers to easily increase their hours of work by small increments. As the stylised facts from Germany and the data showed, many women are working part time. An increase of the school day by a few hours per day might not provide enough childcare to allow women already working to transition from part time to full time work.

Table 32: Estimates on Employment

	(1)	(2)
	ols	logit
VARIABLES	employed	employed
treatment	0.071**	0.756**
	(0.031)	(0.299)
pschildren	-0.053***	-0.467***
	(0.012)	(0.096)
access	-0.002	-0.012
	(0.014)	(0.133)
preschoolkids	-0.324***	-2.334***
	(0.016)	(0.118)
olderkids	-0.041***	-0.360***
	(0.009)	(0.081)
Individual FE	Yes	Yes
State-year dummies	Yes	Yes
Observations	17,234	17,234
Individuals	2,161	2,161
(Pseudo) R-squared	0.656	0.083

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Another explanation could be a positive income effect as a result of this implicit childcare subsidy. As the women in the sample become wealthier as a result of the free childcare, they could be substituting away from labour to leisure. As the theoretical models presented in the Methodology section showed, the sign of the change on hours due to the extension of the school day was unclear. If women have heterogeneous preferences, then some women could increase their hours as a result of the policy change and some women could decrease them. On average this effect will appear as zero, which could explain the coefficients in these regressions as they estimate the average effect on hours.

In Tables 32 and 33, the coefficient on the access variable is small, negative, and not statistically significant. The access variable is a binary variable for whether or not a woman's closest primary school is a full day school. This essentially means that the access variable is picking up the effect of living in a certain geographic region at a certain time. Since it is small and not statistically significant, this means that we do not observe any region-time specific effect associated with the school reform that is influencing labour supply.

Another point to consider when assessing these results are rigidities in the labour market. If there are rigidities that prevent varying hours or extending contracts right away, we might expect delays in how the treatment affects outcomes. It is possible that although women might be made aware of their closest primary school switching over to become a full day facility, they are unable to alter their employment status or current hours worked right away due to the rigidity of contracts. At the same time, schools might not announce their switch-over until shortly before the school

Table 33: Estimates on Hours

VARIABLES	(1) lnweeklyhrs	(2) weeklyhrs
treatment	-0.069	-1.339
	(0.050)	(0.853)
pschildren	-0.104***	-2.689***
	(0.017)	(0.343)
access	-0.012	-0.506
	(0.020)	(0.461)
preschoolkids	-0.407***	-8.830***
	(0.031)	(0.612)
olderkids	-0.050***	-1.299***
	(0.015)	(0.305)
Individual FE	Yes	Yes
State-year dummies	Yes	Yes
Observations	20,985	20,985
Individuals	4,614	4,614
R-squared	0.756	0.784

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

year begins, which would not allow many women the chance to change their labour supply right away in response to the treatment. I check for a delay in the treatment effect by regressing the aforementioned outcome variables on a lagged treatment variable. Here the lagged treatment variable is the interaction of a lagged access variable and the non-lagged variable for having primary school aged children. This is because the lag should arise as a result of delayed access, not delays in becoming a parent.

The results from these regressions are presented in Table 34. Here the results show that the lagged treatment variable has a small coefficient in all of the regressions. In the hours regressions, it is small and negative, and in the employment regressions, it is small and positive. The inclusion of the lagged treatment variable does not change the treatment variable from having a statistically significant effect of increasing the probability of being employed, as estimated by the logit regression. The marginal effect associated with this logit coefficient is 0.072, which is similar to the treatment coefficient calculated in the non-lagged treatment regressions. The introduction of the lagged treatment variable does, however, induce a small, negative, and statistically significant impact on weekly hours. In these results we observe negative coefficients on the lagged treatment variable in the regressions on level and log hours, perhaps indicating an income effect or labour market rigidities that do not allow women to simply extend their working hours by a small, incremental amount at a current job. These coefficients, however, are not statistically different from zero and very small.

The inclusion of the lagged treatment variable in Table 34 does not change the coefficients on the treatment variable significantly from those results presented in Tables 32 and Table 33. The lack

Table 34: Estimates using Lagged Treatment

	(1)	(2)	(3)	(4)
	ols	logit	ols	ols
VARIABLES	employed	employed	lnweeklyhrs	weeklyhrs
treatment	0.053	0.695*	-0.074	-1.627*
	(0.038)	(0.422)	(0.069)	(0.959)
treatmentlag	0.005	0.085	-0.010	-0.080
	(0.043)	(0.530)	(0.064)	(1.060)
pschildren	-0.044***	-0.411***	-0.092***	-2.306***
	(0.013)	(0.111)	(0.019)	(0.371)
accesslag	0.007	0.085	0.012	-0.055
	(0.016)	(0.157)	(0.022)	(0.508)
preschoolkids	-0.327***	-2.422***	-0.403***	-8.604***
	(0.018)	(0.136)	(0.035)	(0.679)
olderkids	-0.010	-0.050	-0.045***	-1.176***
	(0.010)	(0.097)	(0.016)	(0.332)
Individual FE	Yes	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes	Yes
Observations	28,637	13,282	17,248	17,248
Individuals	5,343	1,713	3,686	3,686
(Pseudo) R-squared	0.660	0.084	0.761	0.792

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

of any observable difference here indicates that there is no increased effect on labour supply as a result of being treated one year earlier.[18]

The lagged access variable in Table 34 is never statistically significant, indicating that there is no region-time specific effect of the school reform impacting labour supply. In these specifications, however, the sign on the lagged access variable changes and is primarily positive, albeit very small.

3.6.2 Mothers Only

All of the regressions presented in Tables 32, 33, and 34 include both women with and without primary school aged children. As previously mentioned, this essentially means that there are two control groups for the treated women: those who had access to a full day school, but do not have primary school aged children, and those who have primary school aged children, but did not have access. I have controlled for whether or not a woman has primary school aged children in every regression, as well as whether or not she has access to a full day facility, but the mixture of this control group is something to keep in mind when analysing the results. If we believe that women who have children are fundamentally different than women who do not, then the common time trends assumption may be violated for this portion of the control group and the treated women.

In order to account for this potential issue, I run the same models using the sub-sample of only

[18] I also estimate the same models presented in Table 34, but only include the lagged treatment variable. The coefficients on this variable are similar in sign and magnitude on the treatment variable in Tables 32 and 33, but are not statistically significant. This confirms there is no effect on labour supply from being treated one year earlier.

Table 35: Estimates on Mothers with Primary School Aged Children Only

	(1)	(2)	(3)	(4)
	ols	logit	ols	ols
VARIABLES	employed	employed	lnweeklyhrs	weeklyhrs
access	0.029	0.178	0.020	0.119
	(0.060)	(0.630)	(0.112)	(1.601)
preschoolkids	-0.148***	-1.308***	-0.089	-1.776
	(0.037)	(0.293)	(0.077)	(1.134)
olderkids	0.016	0.135	-0.000	-0.011
	(0.023)	(0.209)	(0.033)	(0.529)
Individual FE	Yes	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes	No
Observations	1,216	1,216	2,179	2,179
Individuals	233	233	751	751
(Pseudo) R-squared	0.727	0.173	0.815	0.854

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

women who have primary school children. Since all of these women have primary school aged children, I no longer need to interact the binary variable for having primary school aged children with the binary variable for having access to a full day school to get my treatment variable. Because of this, the variable of interest is whether or not they had access to a full day school. I also do not need to control for whether or not they have primary school aged children, since they all do.

The results of these regressions may be seen in Table 35.[19] The coefficients on the access variable in Table 35 exhibit similar signs and magnitudes as the previous regressions run on the full sample, but are not statistically significant. The variable for "access" is small and positive in all regressions, including the hours regressions. The marginal effect associated with the reported logit coefficient in Table 35 is 0.020, which again is very similar to the effect size estimated using OLS. The similarity in signs and magnitudes on the employment regressions, but loss of significance may be caused by the much smaller sample size and the loss of power as a result. The sample size has indeed decreased by a significant amount. In these regressions, I am estimating the employment regressions off of a sample of only 233 women as compared to 2,161 women in the full sample regressions. Here I estimate the hours regressions using 751 women and in the full sample I have 4,614 women. This is due in part to the fact that this is an unbalanced panel and I do not always observe women when they have primary school aged children and also to the fact that many women in the sample do not have any children. These decreases in sample size, however, cause me to lose significant power in these regressions.

Overall, these similar results in terms of signs and magnitudes, indicate that the inclusion of women who do not have primary school children in the sample is not driving the results.

[19] I also ran the same employment and hours regressions on a different sub-sample of mothers. This time, I included any woman who has ever been a mother, not just the women I observe at the time they have primary school children. In this specification, the magnitudes and signs of coefficients were similar to the mothers-only regressions presented in this section and for this reason have not been included in this chapter.

Table 36: Heterogeneous Treatment Effects: Single Mothers

	(1)	(2)	(3)
	ols	ols	ols
VARIABLES	employed	lnweeklyhrs	weeklyhrs
treatment	0.061*	-0.090	-1.755*
	(0.036)	(0.056)	(0.937)
singmom	-0.151***	0.024	0.355
	(0.016)	(0.029)	(0.619)
treatment*singmom	0.049	0.116	2.298
	(0.062)	(0.104)	(1.827)
pschildren	-0.052***	-0.104***	-2.696***
	(0.013)	(0.017)	(0.343)
access	-0.003	-0.011	-0.491
	(0.014)	(0.020)	(0.462)
preschoolkids	-0.319***	-0.408***	-8.840***
	(0.016)	(0.031)	(0.611)
olderkids	-0.019**	-0.052***	-1.330***
	(0.009)	(0.015)	(0.307)
Individual FE	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes
Observations	36,149	20,985	20,985
Individuals	6,964	4,614	4,614
R-squared	0.659	0.756	0.784

Clustered standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

3.6.3 Heterogeneous Treatment Effects

I also explore the possibility of heterogeneous treatment effects by interacting the treatment variable with demographic control variables since the effects of the treatment may differ for different women. First, I look at single mothers and then at women whose husband is employed since these two types of women could have fundamentally different responses to being treated.

In Table 36, I present the results from the interaction of the treatment variable with a binary variable for whether or not the woman is a single mother. If we think that single mothers are more likely to work due to economic concerns, then their response to the treatment may differ in a key way.

These results do not show a statistically significant heterogeneous treatment effect for single mothers versus the rest of the sample. Even though the coefficient on the interaction term of treatment and single mothers is positive in all three regressions, it is not statistically different from zero. This indicates that the treatment has not been different for single mothers in this linear framework. Interestingly, in Column (1) the coefficient on the binary variable for single mothers is statistically significant at the one percent level and large. This implies that single mothers are much less likely to be employed than the rest of the women in this sample.

Turning to look at heterogeneous treatment effects for women whose husband is employed versus those whose husband is not working, I find some statistically significant effects. Whether or not the

Table 37: Heterogeneous Treatment Effects: Working Husband

	(1)	(2)	(3)
	ols	ols	ols
VARIABLES	employed	lnweeklyhrs	weeklyhrs
treatment	-0.023	0.099	3.826
	(0.053)	(0.080)	(2.945)
husbandwork	0.029**	0.000	0.154
	(0.013)	(0.018)	(0.385)
treatment*husbandwork	0.088	-0.180**	-5.623*
	(0.056)	(0.086)	(2.919)
pschildren	-0.043***	-0.094***	-2.404***
	(0.013)	(0.019)	(0.377)
access	0.002	0.009	-0.175
	(0.017)	(0.024)	(0.489)
preschoolkids	-0.304***	-0.388***	-8.328***
	(0.018)	(0.033)	(0.644)
olderkids	-0.031***	-0.050***	-1.364***
	(0.010)	(0.017)	(0.336)
Individual FE	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes
Observations	26,712	15,526	15,526
Individuals	5,483	3,634	3,634
R-squared	0.666	0.772	0.806

Clustered standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

husband is employed is an indicator of socio-economic status and economic security of the family. At the same time, it might serve as a predictor of whether or not the woman decides to work.

Table 37 shows the results from interacting the treatment variable with the binary variable for whether or not the woman's husband is employed. In the hours regressions in Columns (2) and (3), I find that the women who are treated whilst their husband is employed decrease their hours. This would indicate that the negative coefficient on the treatment variable in many of the hours regressions is driven by an income effect. These women are potentially more economically secure due to their working partner and can substitute away from labour towards leisure as a result of the implicit childcare subsidy. I do not find any significant heterogeneous effect for these women in the employment regression in Column (1). Overall the evidence for heterogeneous treatment effects along the dimensions explored in this section is weak.

3.7 Robustness

3.7.1 Alternative Clustering

As a robustness check, I run the same models presented in the Results section, but cluster the standard errors differently. The results in Table 38 show that clustering on the school variable did not change the standard errors enough to change the significance test results of any of the coefficients, even though there are fewer schools than individuals in the data set. It seems logical

Table 38: Results with SE's Clustered at the School Level

	(1) ols	(2) logit	(3) ols	(4) ols
VARIABLES	employed	employed	lnweeklyhrs	weeklyhrs
treatment	0.071**	0.756**	-0.069	-1.339
	(0.030)	(0.297)	(0.050)	(0.816)
pschildren	-0.053***	-0.467***	-0.104***	-2.689***
	(0.012)	(0.102)	(0.016)	(0.324)
accessgts	-0.002	-0.012	-0.012	-0.506
	(0.014)	(0.143)	(0.019)	(0.438)
preschoolkids	-0.324***	-2.334***	-0.407***	-8.830***
	(0.016)	(0.133)	(0.028)	(0.556)
olderkids	-0.041***	-0.360***	-0.050***	-1.299***
	(0.008)	(0.082)	(0.014)	(0.303)
Observations	17,223	17,223	20,985	20,985
(Pseudo) R-squared	0.656	0.083	0.756	0.784

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

in this context to cluster at the individual level since labour market decisions are ultimately being made by the individual and this is where correlation in the error term could arise.

Table 39 shows the results of clustering the standard errors along both the person and the time dimension.[20] Cameron et al. (2011) outline the advantages of two-way or multiway clustering, especially in panel contexts. In this case, the two-way clustering changes the statistical significance of the coefficient on log weekly hours. The coefficient is still negative and this decrease amounts to approximately one hour and twenty minutes per week. This result is similar to Contreras et al. (2010), who also found that the extension of the school day in Chile had a negative impact on mothers' hours worked. Contreras et al. (2010) pointed to an income effect as the explanation for this negative effect: as the women in the sample become wealthier as a result of the implicit childcare subsidy, they substitute away from labour towards leisure.

3.7.2 Dropping Teachers

In my previous analysis, I have ignored any possible impacts the extension of the school day may have had on the labour market. This might not be reasonable given how large the reform is and the impact it may have had on the labour market for teachers. Since schools needed to hire many new teachers as a result of extending the school day, there has been increased demand for teachers across Germany.

Statistics from the Federal Statistical Office show that in the 2012-2013 school year, 88 percent of all primary school teachers were women (Statistisches Bundesamt). Since teaching is a traditionally

[20]It should be noted that the coefficients in this specification differ slightly than previous specifications, which is due to the fact that this specification only includes year dummies and not state-year dummies. Because this analysis must be undertaken on the servers at the German Institute for Economic Research (DIW), I have not yet had the chance to run the new specification with state-year dummies.

Table 39: Two-way Clustered Results

	(1)	(2)	(3)
	ols	ols	ols
VARIABLES	employed	lnweeklyhrs	weeklyhrs
treatment	0.072**	-0.070	-1.335*
	(0.032)	(0.043)	(0.749)
pschildren	-0.057***	-0.109***	-2.747***
	(0.013)	(0.018)	(0.417)
accessgts	0.009	0.002	-0.314
	(0.013)	(0.018)	(0.402)
preschoolkids	-0.329***	-0.411***	-8.853***
	(0.017)	(0.024)	(0.473)
olderkids	-0.043***	-0.053***	-1.323***
	(0.008)	(0.014)	(0.326)
Individual FE	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Observations	34,578	19,768	19,768
R-squared	0.049	0.054	0.057
Individuals	5,385	3,391	3,391

Two-way cluster-robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

female dominated career in Germany, the large increase in demand for teachers could affect the mothers in my sample. Teaching is also a career that allows women to combine work with childcare in a relatively straightforward fashion since their working hours do not extend beyond school hours. Additionally, many teachers in Germany also work part time. Of all teachers working in primary schools in the 2012-2013 school year, 38.65 percent of them were employed on a part time basis (Statistisches Bundesamt).

In order to disentangle the increased demand for teachers from the implicit childcare subsidy the mothers receive as a result of the extended school day, I drop all women from my sample who ever worked as teachers.[21] This leaves me with 6,695 women. I then run the same regressions on employment status and hours on this sub-sample. These results may be seen in Table 40.

These results are similar to the results on the full sample in terms of the employment regressions. Columns (1) and (2) in this table show that women who are treated and have never worked as teachers are still approximately 7 percent more likely to be employed. The marginal effect associated with the logit coefficient in Column (2) is 0.069. These results indicate that the increased demand for teachers is not driving the change at the extensive margin.

When we turn our attention to Columns (3) and (4), the results differ from those using the full sample. Now the negative effect of being treated on hours worked, both the log and level, is statistically significant at the 10 percent level; however, the size of the coefficients is similar to those obtained from the full sample. This would reinforce the idea that extending the school day

[21] I drop any woman who has worked as a teacher at any type of school, not just primary, because the reform to extend the school day has also occurred at the secondary schooling level.

Table 40: Estimates Without Teachers

	(1)	(2)	(3)	(4)
	ols	logit	ols	ols
VARIABLES	employed	employed	lnweeklyhrs	weeklyhrs
treatment	0.069**	0.692**	-0.092*	-1.633*
	(0.032)	(0.301)	(0.048)	(0.861)
pschildren	-0.050***	-0.435***	-0.103***	-2.635***
	(0.013)	(0.100)	(0.018)	(0.354)
access	-0.002	-0.014	-0.017	-0.553
	(0.015)	(0.135)	(0.020)	(0.473)
preschoolkids	-0.323***	-2.352***	-0.411***	-8.790***
	(0.017)	(0.123)	(0.032)	(0.630)
olderkids	-0.041***	-0.359***	-0.046***	-1.189***
	(0.010)	(0.082)	(0.015)	(0.310)
Individual FE	Yes	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes	Yes
Observations	16,135	16,135	19,627	19,627
Individuals	2,040	2,040	4,345	4,345
(Pseudo) R-squared	0.659	0.083	0.761	0.792

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

has made mothers who were already working decrease their hours due to the income effect of the implicit childcare subsidy.

3.7.3 Dropping the Women Who Move

One way that women may be able to change their treatment status is by moving house so that their new closest primary school is a full day school. These women would undermine my identification strategy because their assignment to treatment is no longer random. Additionally, if the women who want to move closer to a full day school are women who have a strong preference to work, this would overstate the importance of access to a full day school on mothers with primary school aged children. This is why I run the same participation and hours models on a sub-sample of women that excludes women who have moved house from a home where the closest school was not a full day school to a home where the closest school is a full day school. Although I do not actually know whether or not this is the reason these women have moved (this is not explicitly asked in the SOEP), I still drop these women as an additional robustness check.

I only observe 7 women in the data who move houses and thereby directly change their treatment status. These women all had primary school aged children at the time of their move and move from a school without a full day to a school with a full day. I first drop these 7 women to ensure that my treatment effects are not being driven by these "movers." The results of this analysis may be seen in Table 41. As this table shows, dropping the 7 women who moved house and thereby changed their treatment status does not impact the results. I still find a positive and statistically

Table 41: Estimates Without Women Who Move and Change Treatment Status

	(1)	(2)	(3)	(4)
	ols	logit	ols	ols
VARIABLES	employed	employed	lnweeklyhrs	weeklyhrs
treatment	0.075**	0.790**	-0.064	-1.251
	(0.032)	(0.312)	(0.051)	(0.869)
pschildren	-0.054***	-0.472***	-0.105***	-2.700***
	(0.012)	(0.097)	(0.017)	(0.345)
access	-0.003	-0.019	-0.012	-0.502
	(0.014)	(0.134)	(0.020)	(0.463)
preschoolkids	-0.322***	-2.328***	-0.408***	-8.830***
	(0.016)	(0.119)	(0.031)	(0.617)
olderkids	-0.041***	-0.369***	-0.051***	-1.302***
	(0.009)	(0.081)	(0.015)	(0.307)
Individual FE	Yes	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes	Yes
Observations	17,162	17,162	20,948	20,948
Individuals	2,154	2,154	4,608	4,608
(Pseudo) R-squared	0.656	0.083	0.756	0.784

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

significant impact of the policy at the extensive margin and no effect at the intensive margin.

It is possible that some families decide to move a few years before their children are ready to enter primary school and still choose to move closer to a full day school. I observe a total of 137 women who move house from a home where the closest primary school was not a full day school to a home where the closest school is a full day school. As previously mentioned, I do not know the reason for their move, however, in the interest of robust results, I also drop these women from the sample in case they were anticipating the implicit childcare subsidy in advance and thereby changing their treatment status. These results may be seen in Table 42.

These results are similar to the results obtained on the full sample, although the magnitude of the coefficients on the employment regressions increases when I drop all the women who have moved to be closer to a full day school. The marginal effect associated with the logit coefficient reported in Column (2) is 0.103, which is larger than in the standard regressions on the full sample. Statistically speaking, however, these differences are negligible and reinforce the point that women are not selecting into treatment.

These robustness checks on women moving house in order to select a full day primary school show that this is not driving the overall results I find. The extended school day has enabled mothers to enter the labour market if they were not working before and has not impacted the number of hours they were working.

Table 42: Estimates Without Women Who Move

	(1)	(2)	(3)	(4)
	ols	logit	ols	ols
VARIABLES	employed	employed	lnweeklyhrs	weeklyhrs
treatment	0.094***	1.006***	-0.034	-0.705
	(0.036)	(0.370)	(0.057)	(0.920)
pschildren	-0.055***	-0.483***	-0.101***	-2.601***
	(0.013)	(0.099)	(0.018)	(0.357)
access	-0.012	-0.095	-0.008	-0.426
	(0.017)	(0.165)	(0.023)	(0.511)
preschoolkids	-0.322***	-2.356***	-0.389***	-8.211***
	(0.017)	(0.127)	(0.032)	(0.637)
olderkids	-0.041***	-0.372***	-0.050***	-1.291***
	(0.010)	(0.084)	(0.015)	(0.315)
Individual FE	Yes	Yes	Yes	Yes
State-year dummies	Yes	Yes	Yes	Yes
Observations	15,793	15,793	19,531	19,531
Individuals	2,025	2,025	4,418	4,418
(Pseudo) R-squared	0.664	0.083	0.766	0.796

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.8 Conclusion

The reform to extend the school day in Germany has been one of the largest reforms ever undertaken in their school system. As shown in the descriptive statistics of this chapter, the reform is far from complete as many primary schools still have to switch-over to an extended school day. This entails building cafeterias and hiring new teachers. This lag in the reform has created a unique natural experiment, which allows me to look at how extending school hours affects maternal labour supply in a way few other studies have previously been able to do.

I find robust effects of the extension of the primary school day on maternal labour supply. Mothers of primary school aged children are 7 percent more likely to enter the labour market once they are given access to a full day primary school. This result is robust to changing the sample to include only mothers of primary school aged children and to dropping women who may have moved house in order to live near a full day school. It is a large effect that shows this policy has been successful at drawing mothers into the labour market.

At the intensive margin, I find less robust evidence. In most specifications, the effect of the reform on hours worked is small, negative, and statistically insignificant. In the few specifications where the coefficient is statistically significant, its magnitude is still very small and negative. This result of mothers decreasing their working hours when given an implicit childcare subsidy can be explained by the income effect of the implicit childcare subsidy dominating the substitution effect when leisure is a normal good. The unintended consequence of mothers potentially reducing working hours is something of which policymakers should be aware.

Childcare policies may be used to draw mothers into the labour market after having children or extend their working hours if already working; however, childcare costs and the length of the school day must be taken into consideration by policymakers. In a country such as Germany, where female labour supply is dominated by part time work and stay-at-home mothers, such policies can enact fundamental change to the labour market if implemented appropriately.

4 Non-cognitive Peer Effects

4.1 Introduction

How peers impact each other is of great interest, but unfortunately, not fully understood. From smoking, teen pregnancy, and other risky behaviours, to grades and labour market outcomes, we are interested in identifying the ways in which friends and acquaintances impact each other, for better or for worse. Understanding the mechanisms through which this works and the degree to which peers affect outcomes, however, is not an easy task as the previous literature in this field has shown. Data limitations and difficulties with identification have limited previous research and made policy implications often intangible.

From an education policymaking perspective understanding "peer effects" is important. Peers influence each other in the classroom in terms of the learning environment they help to create. This learning environment is influenced by pupils' cognitive ability, but also by their personality, behaviour, and the non-cognitive traits they possess and can greatly impact an individual's learning outcomes. There is growing evidence that non-cognitive ability may matter just as much or even more for labour market outcomes (Borghans et al., 2008), which should make us think that it matters as much for learning outcomes. As will be discussed in the Literature Review section, most of the literature looking at non-cognitive peer effects has focused on how disruptive behaviour impacts individual learning outcomes. While disruptive behaviour may be correlated with certain non-cognitive traits, there are clearly other dimensions of interest. The lack of evidence on peer effects through the channel of non-cognitive traits means that education policymakers may be missing out on key evidence when creating interventions to improve academic outcomes, especially if non-cognitive traits are malleable.

In this chapter, I use data from Flanders, Belgium to look at the impact of cognitive and non-cognitive peer effects on academic outcomes. The construction of this data set enables me to avoid many of the pitfalls that arise from standard administrative data used in much of the peer effects literature and better identify the effect and narrow down the underlying mechanism. I find that having more conscientious peers in a classroom is positively related to individual math and Dutch performance. In the case of Dutch, however, having a greater variance of peer conscientiousness in a classroom is negatively related to individual test scores. I also find that having more extroverted peers in a classroom is negatively associated with individual math performance. I find no support for peer IQ or past subject performance positively impacting results in the standard linear-in-means framework, but some evidence of non-linear cognitive peer effects in the case of Dutch. Overall, the findings in this chapter show that peers influence each other's learning outcomes in ways beyond the traditional channels of IQ and past subject performance.

4.2 Existing Literature

The relevant literature to this work may be divided into studies that focus on non-cognitive traits and economic outcomes and studies that focus on how peers affect each other, peer effects. Very few studies look explicitly at how the non-cognitive traits of peers impact individual outcomes, but I will highlight those that do. I begin with a discussion of the existing literature on non-cognitive traits and economics and then turn my attention to previous work on peer effects.

4.2.1 Non-cognitive Traits

Over the past 15 years, there has been increasing interest placed at the intersection of non-cognitive traits and economic outcomes. Heckman and Rubenstein (2001) point out that many early economic models of human capital formation, as well as signalling models, completely ignored non-cognitive ability; instead, these models treated ability as something one dimensional, with that dimension being purely cognitive. In these models, the cognitive dimension was measured by either an intelligence quotient (IQ) test score or some other type of academic achievement. Despite this, there is much anecdotal evidence that things like perseverance and other personality traits affect academic and especially labour market outcomes; apart from sheer luck, how else can we explain people with low IQ still succeeding in the labour market while people with high IQ might not be as successful as models would predict? Almlund et al. (2011) and Borghans et al. (2008) provide an exhaustive overview of the previous work on personality psychology and economics that try to address this very question.[22] I will highlight the studies most relevant to this work, which underline the importance of non-cognitive traits in the education context. Most of the studies mentioned here examine how non-cognitive traits impact academic or other outcomes, while a few use non-cognitive measures as outcome variables.

In this chapter, non-cognitive ability is closely related to personality and "soft skills," which Heckman and Kautz (2012) define as "personality traits, goals, motivations, and preferences that are valued in the labour market, in school, and in many other domains." In this chapter, I use measures of soft skills that draw heavily on the "Big Five" taxonomy, a term first coined by Goldberg (1971). As Goldberg (1990) points out, the five factor model (hence the "Big Five") builds off of previous work concerning the classification of personality. His work shows, however, that the thousands of adjectives used to describe personality can be distilled into five orthogonal factors, which are: openness (referred to by some as "culture" or "intellect"), conscientiousness, extraversion, agreeableness, and neuroticism (versus emotional stability) (Goldberg, 1990). Although there

[22] Borghans et al. (2008) make an important distinction regarding non-cognitive traits as compared to much of the economics literature in this area. They purposefully avoid the distinction between "cognitive" and "non-cognitive" traits because ultimately every aspect of human behaviour relies on some degree of cognition. Instead, they focus on personality with a significant amount of focus placed on the Big Five. Since much of the economics literature, especially work by Heckman, uses the division between cognitive and non-cognitive, I choose to do the same. I am aware, however, of the discussion in the literature as to the validity of terms.

is still some debate within the personality psychology literature as to how many orthogonal factors actually exist (with a range of two to five), the Big Five taxonomy is generally well accepted (John and Srivastava, 1999). To provide further information on the Big Five, I reproduce Table 1 from Almlund et al. as Table 43. As Almlund et al. point out, this is a widely used and accepted taxonomy of personality. Many of the studies I cite in this section also draw on the Big Five taxonomy for their measures of non-cognitive traits, although not all. Often, economists are forced to use whatever measures might be available in existing data sets, many of which were not created with psychological measures in mind.

Table 43: The Big Five Domains

Factor	Definition of Factor
I. Openness to Experience (Intellect)	The tendency to be open to new aesthetic, cultural, or intellectual experiences.
II. Conscientiousness	The tendency to be organized, responsible, and hardworking.
III. Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; chracterized by positive affect and socialability.
IV. Agreeableness	The tendency to act in a cooperative, unselfish manner.
V. Neuroticism (Emotional Stability)	Neuroticism is a chronic level of emotional instability and proneness to psychological distress. Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes.

Source: American Psychological Association Dictionary (2007) in Almlund et al. (2011)

Heckman et al. (2013) look at how the well-known Perry Preschool Project affected non-cognitive outcomes. This is one of the few studies in which non-cognitive measures serve as the outcome variable. The evidence that the Perry Preschool Project impacted academic and economic outcomes was already well established. Heckman et al. (2013) point out that although there was a fade out of the effect on IQ, personality was affected in a much more long-term way. Their data includes 43 measures of child personality, which following an explanatory factor analysis (EFA), they codify into three factors: cognition, externalising behaviour, and academic motivation. They find that the intervention affected externalising behaviour in a statistically significant way for all participants, and that both cognition and motivation were positively affected for girls. They are unable to say, however, if the "personality skills" developed by the participants of the programme causally impacted the academic and other outcomes.

This issue of reverse causality between personality measures and outcomes is prevalent in the literature. Does personality drive the outcome or does the attainment of the outcome fundamentally alter personality? Many studies, as in this one, use non-cognitive measures that pre-date the outcome with the hope that this will minimise the issue of reverse causality. Heckman et al. (2006) point out that this strategy might not be enough to avoid biased estimates. If we believe that non-cognitive traits evolve over time, much in the way that skills do, and that they are determined by some other latent factor that also influences the outcome, then measurement error will also be an issue in this context (Heckman et al., 2006).

Because of the issues surrounding measurement error, many economists have turned to latent factor models. While computationally demanding and not applicable to all data sets, latent factor models allow researchers to better understand skill formation over the life cycle and more accurately estimate the returns to skills. Heckman et al. (2006) explicitly model non-cognitive traits as an additional dimension in a latent factor framework using a Roy Model. They conclude that many labour market and other life outcomes can be explained by cognitive and non-cognitive skills.

Heckman and Rubenstein (2001) provide evidence on the importance of non-cognitive skills on GED test outcomes[23] in the United States. They do not identify any one non-cognitive skill in their analysis of high school dropouts, people who complete their GED, and people who graduate high school, but by comparing the groups and controlling for cognitive ability using the Armed Forces Qualification Test (AFQT), they show that GED recipients earn less, have lower hourly wages, and obtain a lower level of schooling than high school dropouts. They point to some unidentified non-cognitive skill as accounting for this discrepancy since they have controlled for cognitive ability.

Bowles et al. (2001) look at how non-cognitive traits impact the productivity of human capital accumulation and thereby affect labour market outcomes. Similarly to Heckman and Rubenstein, Bowles et al. also try to explain the observed wage differences between seemingly similar individuals. They cite the literature linking physical appearance to earnings to point out that other research has already identified non-cognitive sources of wage variation. They develop a model that includes both cognitive and non-cognitive factors as determinants of earnings. One key difference between this work and many of the others cited here is their focus on "situation specific behaviour," how an individual acts in certain circumstances, as opposed to underlying traits, which are better measured using personality tests. People adapt their behaviour to given contexts, which is not necessarily a true indication of their underlying characteristics. There was a significant debate in the literature between behavioural economists, who focused on situation specific behaviour, versus psychologists and economists interested in personality (see Almlund et al. (2011) for an overview of this debate). For the purposes of this work, the focus will be placed on personality, which has been shown to have stable components over time (Borghans et al., 2008).

[23] The "GED" or "General Educational Development" test is a high school diploma equivalency exam.

4.2.2 Peer Effects

In his Handbook Chapter, Sacerdote (2001) uses a "broad definition of peer effects to encompass nearly any externality in which peers' backgrounds, current behaviour, or outcomes affect an outcome. By limiting peer effects to externalities, market-based or price-based effects are excluded" (Sacerdote, 2001). Here we see that peer effects may include aspects of socio-economic status, how a pupil acts in the classroom, or how the pupil's peers perform throughout the school year. Because this definition is so broad, it allows for many different aspects of student background and ability to enter into the peer effect through a variety of mechanisms. For example, how teaching style or level is affected as a result of the ability distribution or behaviour in a given classroom is part of a peer effect. A peer effect can also arise because pupils are very motivated by one or a few very bright pupils in the classroom and want to compete or simply keep up. Similarly, how family background shapes the ability that pupils bring to the classroom and thereby affects how their peers learn is also part of a peer effect. Educated and motivated parents who lobby a school to improve teaching quality are considered a peer effect under this definition as well. Here I use a similarly broad definition as I am interested in how classmates' backgrounds and underlying abilities, both cognitive and non-cognitive, impact their peers' outcomes. I use the term "non-cognitive peer effect" to mean the impact of classmates' underlying non-cognitive ability on an individual.

Much of the previous work on peer effects in education has focused on how the academic performance of peers impacts individual outcomes. Within this body of work, one general division may be made between studies that exploit exogenous variation in how peer groups are formed and studies that attempt to control for selection into peer groups. As this chapter fits into the latter group, these types of papers will be discussed more extensively.[24]

In a paper examining peer effects in North Carolina schools, Vigdor and Nechyba (2007) look at peer effects at the classroom level. They use school fixed effects to control for the selection problem in peer group formation since they use administrative data from North Carolina where no random assignment to schools or classrooms has occurred. The inclusion of school fixed effects in their models does not affect the coefficients on the peer effects variables. They find a significant and positive impact of 3rd grade peer academic performance on 5th grade outcomes. Interestingly, they also look at ability dispersion and find that greater variance in ability within a classroom predicts better math outcomes. I will similarly look at ability dispersion as well as the effect of average ability.

Burke and Sass (2013) use panel data on Florida schools to estimate peer effects at the class and grade level within a school. They directly address the issues of simultaneity, selection, and measurement error in their various specifications. They do not find any indication of peer effects

[24] For a review of papers exploiting exogenous variation in peer group formation see Sacerdote (2001).

when they estimate the standard linear-in-means model, but when they turn to non-linear models, they find significant results at the class level, but not at the grade level. This prompts me to also examine non-linear models of peer effects, something I will discuss in my Empirical Strategy section.

In a cross-European study using data from the Progress in International Reading Literacy Study (PIRLS), Ammermueller and Pischke (2009) also examine the impact of peer ability at the classroom level. Their identification strategy relies on intra-school variation at the class level, arguing that at the primary school level classes within a school are more or less formed on a random basis. Since the PIRLS data includes multiple classes within a grade at one school, they are able to exploit this variation. They find moderately large peer effects, even when accounting for issues such as measurement error, and also find that selection into schools does not drive their results, which is similar to the results obtained by Vigdor and Nechyba in the United States. Their strategy of looking at classes within the same school is similar to the one I will pursue in this chapter.

Betts and Zau (2004) present evidence from the San Diego Unified School District on how previous academic performance of peers affects current school year performance at the class level. Their models include student fixed effects to control for unobserved heterogeneity and find positive and significant effects of peer effects on gains in learning. Much like Burke and Sass, they only find these effects at the class level, not the grade level, which further reinforces the point that something is happening at the actual level of learning, not just amongst a cohort in a school. Because they use administrative data, they do not have any non-academic outcome measures nor do they have any measures of non-cognitive peer effects.

Hoxby (2000) exploits changes in the racial and gender breakdowns of classrooms in the same school in consecutive years to identify peer effects and control for the selection problem. She finds that peer effects are more pronounced intra-race and that a higher percentage of girls in a classroom leads to better learning outcomes; these results also show that previous academic achievement of peers does influence individual learning outcomes. Similarly, Geay, McNally, and Telhaj (2013) use the English National Pupil Database to look at the effect of having more non-native English speakers in a classroom on native speakers' academic outcomes at the end of primary school. Although there is a negative raw correlation between a higher proportion of non-native speakers and lower academic outcomes for native speakers, once they implement their identification strategy and control for selection, this effect disappears.

Neidell and Waldfogel (2008) look explicitly at non-cognitive peer effects and non-cognitive outcome measures in their paper examining the effect of peer group kindergarten attendance on first grade learning outcomes. They deal with the selection problem by including school fixed effects and extensive controls for unobserved pupil and teacher level heterogeneity. Their findings show that high peer externalising problems, behaviours associated with "defiance, impulsivity, disruptiveness, aggression, antisocial features, and overactivity" (Hinshaw, 1992), negatively affect learning outcomes.

Gibbons and Telhaj (2015) also use the National Pupil Database to look at peer effects in the transition from primary to secondary school in England. Due to the large amount of data at their disposal, they are able to aggregate thousands of individual level observations and look at primary school cohorts over time. They use variation in peer group as a result of selecting into a new school to identify peer effects at this cohort level. They find that the quality of peers upon entrance to secondary school, as measured by past performance, positively impacts secondary school test scores. Gibbons and Telhaj is relevant to this work as they also look at primary to secondary school transition, but their data allows them to pursue an identification strategy not possible here.

There has been little work within the Economics literature done using the LOSO data set, which I use in this chapter. The research team who led the LOSO project published several articles on schools, classes, and non-cognitive outcomes. Van Landeghem et al. (2002) use the LOSO data to look at the effect schools and classes have on several non-cognitive measures such as social integration. In this paper, they are also concerned with school effectiveness and how this impacts non-cognitive outcomes. They use characteristics of pupils, classrooms, and schools, such as the gender breakdown in a class, to examine the impact on a variety of non-cognitive outcomes measured at the end of the first two years of secondary school. Their identification strategy does not allow them to make causal inference on any of these relationships.

Fraine et al. (2003) and Opdenakker et al. (2002) each use the LOSO data to look at the impact of classes and schools on Dutch and math achievement respectively. Opdenakker et al. find that the proportion of girls in a class is positively related to academic performance in math. They also find that high initial cognitive ability predicts even higher math outcomes for classes that have a high cognitive ability. They add measures of school and teacher characteristics to control for unobserved heterogeneity at those levels, but their identification strategy does not fully account for unobserved heterogeneity, making causal inference difficult.

Fraine et al. employ a similar methodology to look at academic achievement in Dutch. They find that class composition is very important in predicting language outcomes. Similarly to Opdenakker et al., they find that students in classes with higher initial cognitive ability or a higher proportion of girls have higher Dutch grades at the end of the school year. I build on both of these studies by explicitly modelling initial non-cognitive peer effects at the class level in addition to initial cognitive ability. I also more completely control for selection into class and school.

The results of many of these studies show that having a higher proportion of girls in a classroom positively impacts learning outcomes. The mechanism through which this works, however, has yet to be fully explained. One potential explanation is that girls and boys have on average different strengths and weaknesses in terms of non-cognitive skills. Bertrand and Pan (2011) review the medical and psychological evidence that points to this, e.g. the fact that boys are statistically more likely to be diagnosed with attention deficit hyperactivity disorder (ADHD). Aizer (2008) explicitly looks at ADHD and exploits exogeneity in access to treatment as a way to change peer behaviour. She finds that improvements in peer behaviour improve academic outcomes. Her findings indicate that peer behaviour is perhaps just as important as peer ability in determining outcomes.

Few studies have explicitly modelled the effect of non-cognitive peer effects, apart from focusing on disruptive behaviour, on learning outcomes. This is mostly because few data sets include these types of measures in real detail. I will build on the existing literature in the field to directly assess the importance of non-cognitive peer effects for learning outcomes.

4.3 The Policy Implications of Peer Effects

Traditionally, peer effects have mattered to education policymakers because there is evidence that classroom composition affects outcomes and classroom composition is something they can control. Tracking or streaming policies are intended to group pupils of similar ability together. Most often, ability is measured through previous grades or teacher recommendations. The idea behind these policies is that academically high-performing students benefit from being in the same class as other academically high-performing students; there are no low-performing students bringing down the content level or slowing down the pace of learning. Teachers are able to teach to one similar level, further improving the learning experience. Even if all average or low ability pupils are grouped together, teachers can still teach to one level.

Such policies may have other unintended consequences; academically more able pupils may benefit whilst more average pupils may lose out. For example, children of low socio-economic status end up only being placed with other children of low socio-economic status because of the low existing ability they bring to the classroom. As a result, they may only end up being dragged further down. This phenomenon, which can also work in the opposite direction of high ability pupils improving over time as a result of consistently being tracked with other high ability pupils, is described as the "dynamic tracking problem" (Rothstein, 2010).

Policymakers should also be aware of the general equilibrium effects of tracking or streaming pupils. As Hoxby (2000) points out, the focus in the literature on the linear-in-means model leads "peer effects [to] have distributional consequences but no efficiency consequences." This means that swapping a high achieving student with a low achieving student will have zero overall effect on learning outcomes.

Beyond the traditional tracking or streaming arguments associated with peer effects, there is evidence that peer level behaviour impacts learning outcomes. Much of this literature tends to focus on disruptive behaviour, but is nevertheless revealing about the influence of peer level non-cognitive measures on the individual. In this chapter, I find that peer level non-cognitive traits also impact individual learning outcomes. If this is the case, schools can implement policies to discourage bad behaviour and provide training in positive skill areas. As Heckman et al. (2013) point out, skills can be developed through interventions and are perhaps more malleable than cognitive ability in a long-term way. In the United Kingdom, for example, the Office for Standards in Education, Children's Services and Skills (Ofsted), which rates schools, is particularly interested in classroom behaviour and how it is related to learning outcomes (Ofsted Website). Ofsted has not only placed a focus on "low level disruption" in the classroom, but also included key behaviourial and non-cognitive terms (e.g. "confident", "self-disciplined", and "self-assured") in their grade descriptors (Ofsted School Inspection Handbook, 2015). Interventions to specifically target and develop non-cognitive skills may not only benefit the individual pupil in her life, but have a knock-on effect in terms of positively impacting her peers through peer effects.

4.4 Education in Flanders

Children in Flanders attend compulsory education from age 6 until age 18. Secondary school, the focus of this chapter, begins at age 12. At this point, children pursue one of two streams, known as the A or B stream; in 2010, 84 percent of children were enrolled in the A stream, which has been relatively stable since the school reform of 1989 (OECD, 2011b). The A stream is the standard stream, while the B stream is focused on children with developmental or learning difficulties. There is less differentiation between schools at this point, as pupils still follow a more standardised curriculum with fewer electives and specialization; however, there is a high degree of school autonomy in Flanders (OECD, 2011b). This means that selection into a specific school is more likely since schools can differentiate themselves and thereby attract prospective pupils. After completing the A or B stream, which takes two years, pupils then move to one of four types of upper secondary schools. These four types of upper secondary schools are: general secondary education (ASO, Algemeen secundair onderwijs); technical secondary education (TSO, Technisch secundair onderwijs); secondary arts education (KSO, Kunst secundair onderwijs); and vocational secondary education (BSO, Beroepssecundair onderwijs) (OECD, 2011b). As indicated by their names, these schools range from academically focused to vocationally focused. Pupils can only attend university if they receive a diploma from an ASO or TSO secondary school. In this chapter I only look at the first year of secondary school because of the relative homogeneity between content at this level.

One main feature of the education system in Flanders is the prevalence of Catholic schools, which contributes to the high degree of school autonomy. The Belgian Constitution includes a "freedom to education" clause, which gives any individual the right to open up a school (OECD, 2011b). In order for schools to receive public funding and issue diplomas, however, they must follow a core curriculum set out by the Flemish authorities and they must agree to be inspected by the Flemish authorities (OECD, 2011b). This same "freedom to education" clause in the constitution gives parents the right to school choice, which naturally complicates the identification strategy in this context.

The quality of education in the Flemish context has been deemed relatively high. On every Programme for International Student Achievement (PISA) test including the first round in 2000, pupils in Flanders have consistently performed well above average (OECD, 2011b). Over 80 percent of Flemish adults aged 25-34 years old have completed upper secondary education and in 2008, 42 percent of 25-34 years olds held a tertiary degree, which is higher than the OECD average (OECD, 2011b). Results from the 2009 PISA for all of Belgium showed that are still vast differences in performance based on socio-economic status of pupils and schools. These same results also pointed to significantly lower performance on the part of immigrant children as opposed to native Dutch speakers; the proportion of low-performing immigrant pupils was three times as large as the proportion of low-performing native Dutch speakers (OECD, 2011b). While these stylized facts come from data collected significantly after the period I look at in this chapter, immigration is not new to Flanders and neither are issues of inequality.

4.5 Data and Descriptive Statistics

The data used in this chapter comes from the "Longitudinaal Onderzoek in het Secundair Onderwijs" or in English, "Longitudinal Research in Secondary Education" Project (also known as the LOSO Project). Van Damme et al. (2002) provide an overview of the project and its aim to assess the effectiveness of secondary schooling in Flanders, Belgium. The project began in 1990 with data collection on more than 6,000 secondary school pupils, who were all twelve years old at the time and about to begin secondary school.

These pupils comprise an entire cohort beginning secondary school in 1990, and attending 57 different secondary schools[25] from all regions of Flanders (Van Damme et al., 2002). For each of the 57 schools sampled, the entire cohort of first year pupils (6,411 pupils) is surveyed. There were some pupils at these 57 schools who had to repeat the first year of secondary school for the second time in 1990. These approximately 350 pupils are also included in my analysis, although I take into consideration the fact that they are one year older than the rest of the cohort and have already studied the curriculum, even if they might be lower performing on average.

In subsequent years some pupils had to repeat a grade and therefore ended up in a classroom with entirely new peers. The LOSO project then included some information on these new peers in the data set because the researchers were still interested in following all original pupils. This means

[25]It should be noted that not every secondary school in Flanders is part of this study.

Table 44: Variable Descriptions

Variable Name	**Description**
pupil_id	Unique pupil identifier
school_id	Unique school identifier
dutch_class	Unique Dutch class identifier
math_class	Unique math class identifier
dutch_irtscore_t1	IRT score in Dutch at the end of year 1
math_irtscore_t1	IRT score in math at the end of year 1
dutch_irtscore_t0	IRT score in Dutch at the beginning of year 1
math_irtscore_t0	IRT score in math at the beginning of year 1
getlov_iq	Overall score on Getlov IQ test at beginning of year 1
CON	Teacher-reported score of pupil's conscientiousness at beginning of year 1
AGREE	Teacher-reported score of pupil's agreeableness at beginning of year 1
EXTRA	Teacher-reported score of pupil's extraversion at beginning of year 1
male	Variable for whether or not pupil is a boy
father_teredu	Variable for whether or not pupil's father attended tertiary education
foreign_parent	Variable for whether or not the pupil has at least one foreign born parent
income	Categorical variable for net monthly family income with six categories

that the sample size grew over time. For the purposes of my initial analysis this is not a problem. I only look at the first year of secondary school, during which time all 6,775 pupils are included in the data. The LOSO data is not an administrative data set, but includes nearly all pupils in Flanders of one secondary school cohort.

One characteristic of the secondary schooling landscape in Flanders is the prevalence of private schools. Many of these private schools are religious, and of the 57 schools surveyed in the LOSO Project, 38 are private and 19 are public (Van Damme et al., 2002). The division between private and public schools means that school choice is definitely occurring; as this is guaranteed by the Belgian Constitution, this is not surprising. I will take this into account in my empirical analysis through the inclusion of school fixed effects.

At the beginning of the LOSO Project, all pupils took the Getlov battery for intelligence, an IQ test (Lancksweerdt, 1991), and school achievement tests in Dutch[26] and math. These achievement tests were created by the LOSO project team and thus standardised across all schools. Their parents also filled out a background questionnaire at the beginning of the study, which covered parental education and other family characteristics. The pupils' teachers had to assess their pupils on non-cognitive characteristics and the pupils themselves filled out a questionnaire assessing their motivation and fear of failure. These measures will be used to control for cognitive and non-cognitive ability, as well as socio-economic status. They will also be used to construct the peer

[26]Despite the common perception that "Flemish" is the official language of Flanders, it is actually Dutch (BBC). There are four distinct dialects of Dutch spoken in Flanders, however, the Dutch taught in secondary schools is the standard Dutch.

Table 45: Summary Statistics

Variable Name	**N**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
pupil_id	6,775				
school_id	57				
dutch_class	354				
math_class	357				
dutch_irtscore_t1	6,499	0.160	1.139	-8	3.735
math_irtscore_t1	6,437	0.064	1.309	-8	3.101
dutch_irtscore_t0	6,554	0.019	1.085	-8	5.414
math_irtscore_t0	6,574	-0.032	1.190	-8	5.553
getlov_iq	6,535	100.116	14.917	40.267	142.641
CON	4,417	10.842	3.350	2	15
AGREE	4,416	12.125	2.990	2	15
EXTRA	4,416	11.582	2.746	2	15
male	6,775	0.514	0.500	0	1
father_teredu	6,775	0.197	0.398	0	1
foreign_parent	6,312	0.084	0.277	0	1
income	5,283	3.794	1.289	1	6
1.income	5,283	0.013	0.113	0	1
2.income	5,283	0.152	0.359	0	1
3.income	5,283	0.290	0.454	0	1
4.income	5,283	0.254	0.435	0	1
5.income	5,283	0.159	0.366	0	1
6.income	5,283	0.133	0.340	0	1

effect measures at the classroom level.

At the end of the first year of secondary school, all students were assessed in Dutch and math, again using a standardised assessment created by the LOSO project team. This will serve as my outcome measure, as I wish to see how academic achievement is affected by peers. The inclusion of both cognitive and non-cognitive peer effect terms will help me better understand the channels through which peer effects work. The data set includes detailed measures of cognitive and non-cognitive ability in addition to academic outcomes, such as grades. This level of detail is generally not available in administrative data and the lack of such information on underlying cognitive and non-cognitive ability creates omitted variable bias, a problem in many similar studies looking at peer effects. This will be less of a concern in this work, as such measures are available.

The advantage of looking at the first year of secondary school is that almost all pupils will undertake a similar curriculum, known as the "A stream." There is limited choice in this first year of secondary schooling, so the curriculum is more standardised across and within schools. Streaming has also not yet begun in Flanders at this point in secondary school, so pupils will not be separated into classes based on ability. This means that for many pupils, the peers in their Dutch and math classes will be somewhat random since the curriculum is the same across and within schools and many students will be attending a new school for secondary school with new classmates. Of course there will be some degree of school selection, as families choose schools based on specific characteristics of a school, but the inclusion of school fixed effects will help control for this.

My primary variables of interest are the teacher reported non-cognitive scores for each pupil and the corresponding non-cognitive peer effect terms on the three Big Five measures of agreeableness, conscientiousness, and extraversion. Each pupil's teacher from their primary school was asked to assess the non-cognitive ability of the pupil at the beginning of secondary school. I then use this assessment and an exploratory factor analysis to construct the non-cognitive measures and then validate them.[27] One issue with these measures is the large number of missing values. This has to do with the inability of the LOSO researchers to reach all primary school teachers for each pupil. Since the primary schools did not participate in the LOSO project, there was less scope to track down every primary school teacher. These missing values are most likely not completely missing at random and this should be kept in mind when interpreting the results obtained through this analysis.

Turning to academic performance, I look at the results from the Dutch and math assessments administered at the end of the first year of secondary school. These measures are "Item Response Theory" (IRT) scores and have been calculated by the LOSO researchers in order to make comparison and analysis across assessments possible. This type of measure is preferred in the educational assessments literature due to the fact that it takes the difficulty and discrimination of the test

[27]See the Appendix for a more detailed description of how these measures were calculated and validated.

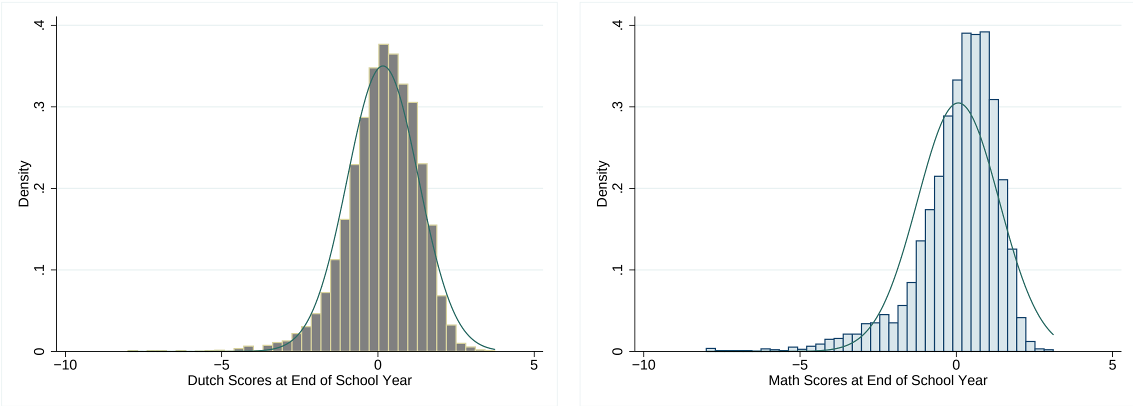


Figure 23: Distribution of Scores at t=1

question into account (DeMars, 2010). IRT scores for the LOSO project were calculated by their research team from raw test scores using a one parameter logistic model commonly known as the Rasch Model. The scores are calculated in such a way that the mean is approximately zero, which implies that having a negative score is below average and a positive score above average. These IRT scores will serve as my outcome variable in all regressions. The assessments were devised by the LOSO researchers, as there is no standardised testing in Flanders. This means that all pupils at all schools participating in the study took the same Dutch and math assessments. Figure 23 shows the distribution of IRT scores in Dutch and math at the end of the first year of secondary school.

I also look at mean academic performance for girls versus boys and immigrant versus non-immigrant children. Table 46 shows mean mathematics performance at the beginning of secondary school (t=0) and at the end of secondary school (t=1) for boys and girls and children with a foreign born parent and children with two Flemish parents. These summary statistics show a higher average performance for children with non-immigrant parents versus those with immigrant parents and slightly better performance for boys over girls. A simple t-test on the means for boys versus girls shows that they are statistically identical, however, a t-test on the means for children with an immigrant versus non-immigrant parent shows a statistically significant difference between their mean math performances.

In Figure 24, I present the distribution of family income for these pupils by category. The variable for income is coded into six categories where "1" means the family income is less than 25,000 Francs[28] per month; "2" between 25,000-40,000 Francs; "3" between 40,000-60,000 Francs; "4" between 60,000-80,000 Francs; "5" between 80,000-100,000 Francs; and "6" more than 100,000 Francs per month. This is the net family income and includes pensions, unemployment and other

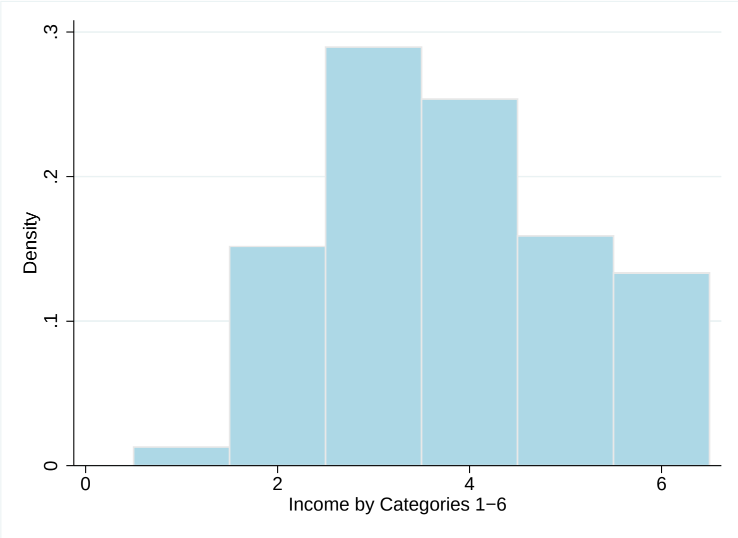
[28]When this data was collected in 1990, the Belgian Franc was still in use. Statistics from the St. Louis Federal Reserve indicate that in September 1990, one U.S. Dollar was worth 32.28 Francs. This means a net income of 100,000 Francs per month in 1990 was equivalent to 3,097.89 USD.

Table 46: Math Performance by Demographic Groups

	Math Performance at t=0			
	Mean	N	Standard Deviation	t-statistic
Immigrant Parent				
Has an immigrant parent	-1.250	489	1.482	
Does not have an immigrant parent	0.060	5,695	1.095	
				24.585
				(0.000)
Gender				
Girls	-0.060	3,203	1.222	
Boys	-0.006	3,371	1.158	
				-1.873
				(0.061)

	Math Performance at t=1			
	Mean	N	Standard Deviation	t-statistic
Immigrant Parent				
Has an immigrant parent	-1.060	493	1.464	
Does not have an immigrant parent	0.158	5,565	1.240	
				20.576
				(0.000)
Gender				
Girls	0.100	3,136	1.300	
Boys	0.035	3,301	1.321	
				1.845
				(0.065)

NB: p-values reported in parentheses below t-statistics



Income Category Key

1. less than 25,000 Francs per month
2. between 25,000-40,000 Francs per month
3. between 40,000-60,000 Francs per month
4. between 60,000-80,000 Francs per month
5. between 80,000-100,000 Francs per month
6. more than 100,000 Francs per month

Net monthly family income

Figure 24: Distribution of Income Categories

benefits, deductions for taxes, and an allowance for children. As Figure 24 shows, the distribution of income by category seems fairly standard. I will use income in some of my analysis to control for the socio-economic status of the individual and her peers.

As discussed in the Literature Review, there is some evidence on the relationship between non-cognitive traits and academic outcomes. In Table 47, I present the correlations between the three non-cognitive measures, conscientiousness, agreeableness, and extraversion, and the overall IQ test score. This table shows that the three non-cognitive measures are relatively weakly correlated with one another and that conscientiousness is most correlated with IQ. This correlation between

Table 47: Correlations between Non-cognitive Measures and IQ

	Conscientiousness	**Agreeableness**	**Extraversion**	**IQ**
Conscientiousness	1			
Agreeableness	0.447	1		
Extraversion	0.535	0.350	1	
IQ	0.644	0.232	0.280	1

Table 48: Summary Statistics at the School Level

Variable Name	N	Mean	Standard Deviation	Minimum	Maximum
Across all schools					
grade_size	57	118.860	76.869	17	363
dutch_class_size	57	18.988	4.034	7.296	26.135
math_class_size	57	18.766	4.099	7.296	26.135
dutch_class_number	57	6.211	3.390	2	18
math_class_number	57	6.263	3.368	2	18
variance math_irtscore_t0	57	1.189	1.004	0.250	5.476
variance dutch_irtscore_t0	57	0.873	0.660	0.177	4.177
variance iq	57	153.452	60.418	53.979	300.060
variance CON	57	8.814	3.063	2.194	16.726
variance AGREE	57	9.070	2.912	3.404	15.944
variance EXTRA	57	7.190	1.857	2	12.048

conscientiousness and IQ has been found in other studies (Borghans et al., 2008).

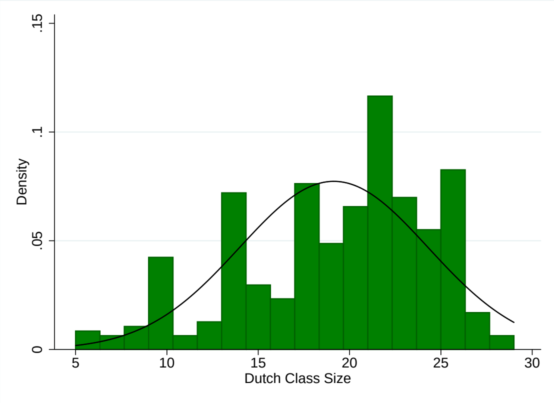
Because the peer effects I am interested in occur at the school and class level, I also present some key descriptive statistics for these levels. In Tables 48 and 49, I present information on class averages for the key cognitive and non-cognitive measures of interest. I also include information on the variances of these variables across classes since I use include the second moments of the peer effect terms in the variance model. Table 49 shows that there is enough variation in the variances of the peer effects terms to pursue this strategy.

Table 48 also presents information on the size of each cohort beginning secondary school at each of the 57 schools, along with the average number of pupils in each math and Dutch class at each school. This table also includes information on the number of math and Dutch classes at each school. It is important for my identification strategy that I have at least two classes at every school since I am including school fixed effects; any school with just one Dutch or math class will be dropped from my estimation as a result. As these descriptives show, however, all of my schools have at least two Dutch and math classes, which means I am able to include all schools from the LOSO project in my analysis.

Figure 25 shows the distribution of the class size for math and Dutch classes across my sample. Both diagrams show more density for larger values of class size.

Table 49: Summary Statistics at the Class Level

Variable Name	N	Mean	Standard Deviation	Minimum	Maximum
Across all classes					
dutch_class_size	354	19.138	5.160	5	29
math_class_size	357	18.978	5.233	5	29
mean math_irtscore_t0	357	-0.218	1.036	-4.025	1.604
mean dutch_irtscore_t0	353	-0.149	0.952	-3.752	1.680
mean iq_m	357	97.953	12.751	60.203	123.551
mean iq_d	354	97.903	12.769	60.203	123.551
mean CON_m	355	10.258	2.696	4	14.688
mean CON_d	353	10.253	2.692	4	14.688
mean AGREE_m	355	11.927	1.449	6.625	14.75
mean AGREE_d	353	11.926	1.450	6.625	14.75
mean EXTRA_m	355	11.398	1.440	5	15
mean EXTRA_d	353	11.399	1.442	5	15
mean male_m	357	0.520	0.342	0	1
mean male_d	354	0.520	0.342	0	1
variance math_irtscore_t0	357	0.766	1.106	0.038	8.397
variance dutch_irtscore_t0	353	0.540	0.752	0.011	7.147
variance iq_m	357	99.123	48.578	17.911	396.868
variance iq_d	354	99.546	48.588	17.911	396.868
variance CON_m	355	6.102	4.077	0	32
variance CON_d	353	6.140	4.089	0	32
variance AGREE_m	355	8.492	4.862	0.25	28.691
variance AGREE_d	353	8.497	4.822	0.25	28.691
variance EXTRA_m	355	6.593	3.344	0	21.929
variance EXTRA_d	353	6.617	3.335	0	21.929



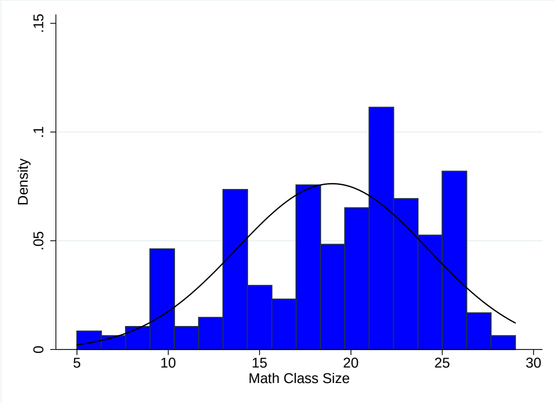


Figure 25: Distribution of Class Size

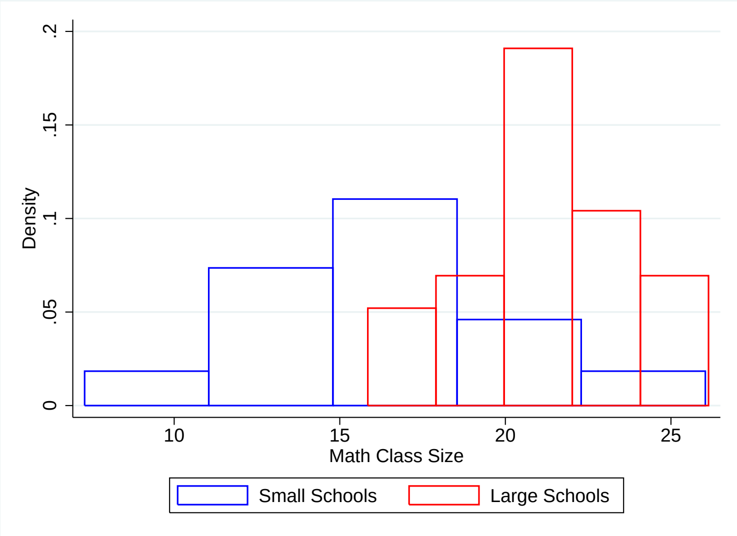


Figure 26: Distribution of Math Class Size by Grade Size

In Figure 26, I present the histogram of average math class size at the school level for small and large[29] schools. This diagram shows that most small classes are found in the smaller schools whilst most large classes are found in the larger schools. I will control for class size in all specifications and include school fixed effects, which should control for any unobserved heterogeneity between schools and some observable heterogeneity between classes.

4.6 Empirical Strategy

Sacerdote (2011) points out in his Handbook Chapter that most of the peer effects literature focuses on the "linear-in-means model." This type of model includes average peer group ability and average peer group characteristics in addition to individual level variables. The drawback of this model is that it forces all peer effects to enter the model linearly and have the same homogenous impact on

[29] Here small schools are all schools below the median grade size of 109 and large schools are all those with a grade size larger than 109.

all students, regardless of where they might fall in the ability distribution (Sacerdote, 2011).

Regardless of the type of model used, there are several key empirical challenges that all researchers looking at peer effects must address. These are: simultaneity, selection, and measurement error. Each of issues poses a challenge to the identification of peer effects and must somehow be taken into consideration by the researcher in the methods used.

Manski (1993) was the first to raise the issue of simultaneity, which many studies do not fully address, as something he termed the "reflection problem." He uses this term to describe the issue that since peers form a group, any group level outcome will be impacted by all members. This means that using peers' smoking habits to estimate the probability that an individual will smoke is inherently biased by the issue that that individual may also impact her peers' smoking habits. Since peers affect each other, their habits and outcomes evolve in a dependent way over time, which leads to a simultaneity bias. In this sense, Manski points out that most peer effects are actually endogenous.

This is an important concern to keep in mind when estimating peer effects in this context, and something I will take into account in my empirical strategy. All of the "peer effect" terms that I include are averages across all pupils in a class, excluding a given individual. This means that every pupil will have a slightly different peer effect term since she is not included in that average. Manski points out that it might not be enough to simply exclude the individual from the peer effect term since this person still influenced her peers. I attempt to avoid this issue by using peer effects measured at the beginning of the school year, at a new school before pupils have had time to influence each other.

In order to explore how many peers a pupil already knew in her class from primary school, I look at the relationship between primary school attendance and secondary school class composition. Since I also know which primary school all of the pupils in the LOSO project attended, I am able to construct a measure of how many peers in a given class went to the same primary school as an individual. This measure tells me what proportion of her peers in a Dutch or math class a pupil knows when starting secondary school. Because of issues associated with the reflection problem, a low proportion of peers from primary school makes my identification strategy cleaner. Figure 27 shows that the median pupil, marked in the graph with a red diamond, knows less than 10 percent of her classroom peers from primary school. This figure also shows that even the pupil at the 75th percentile of the distribution knows less than half of her peers in either Dutch or math from primary school. This gives me confidence that the reflection problem does not pose a major challenge to identification.

The issue of selection still proves challenging for identification. Here the problem arises because students are not randomly assigned to schools (or potentially even classes) and this leads to correlated effects. This means that students in a particular school may have some other common

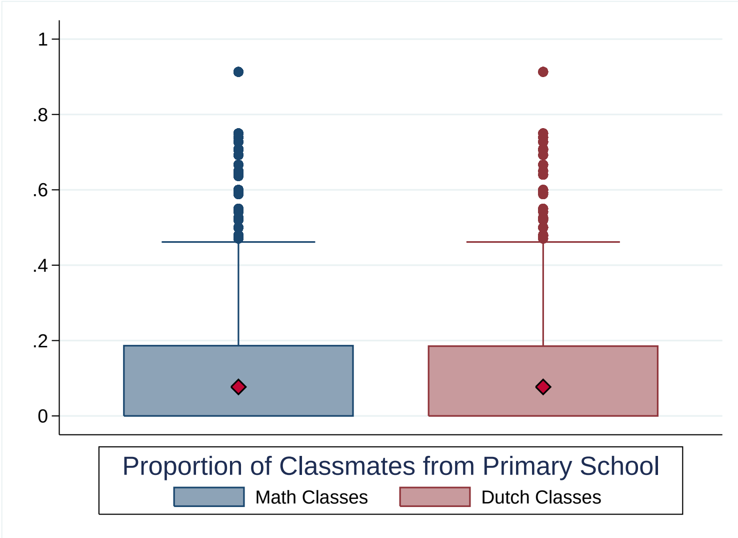


Figure 27: Boxplot of Proportion of Primary School Peers in Secondary School Class

characteristics, such as socio-economic status or parental education levels, that are actually driving the direction of the peer effects. In general most research on peer effects may be divided into studies that exploit exogenous variation in assignment to schools or classrooms and studies that attempt to control for selection through the introduction of school, class, or individual fixed effects. Some studies have also relied on variation in gender or ethnic composition of a particular school in consecutive years (e.g. Hoxby, 2000). In this chapter, I rely on school fixed effects to account for selection. These fixed effects will absorb any time invariant characteristics of the schools, e.g. facilities, demographic characteristics of the student body, teaching style, etcetera. In practice this means that my identification relies on within school variation. If all pupils at a given school are very similar, which may be the case when pupils select into schools, this limits the strength of my identification strategy. As shown in Table 48, however, there seems to be enough variation in the variation of the key variables of interest within the 57 schools that this does not pose a major challenge to identification.

Burke and Sass (2013) point out that fixed effects will not eliminate issues of dynamic tracking, the idea that low-performing students are placed all together in a class the following year and this causes them to continue to perform poorly. This is less of an issue here because I am looking at the first year of secondary school, where all students are new to a school. It is possible, however, that some sort of tracking is happening within schools at the class formation level (if administrators form classes on the basis of primary school grades), so this is something I need to take into account.[30] In general, the problem of dynamic tracking causes a downward bias in the estimation of peer effects due to mean reversion (Betts and Zau, 2004).

[30]Rothstein (2010) provides falsification tests to test for dynamic tracking, but I am unable to do this in this framework as I do not have a panel of these students from primary school onwards. Since tracking officially begins after the first two years of secondary school in Flanders, it does not make sense for me to run these types of tests on the data I do have, but currently do not use.

There is still the issue of how classes at secondary school are formed. If classes are formed randomly, which is usually not the case, then we do not have to consider what might be driving formation. Since they are generally formed with some criteria in mind, however, we need to be concerned with selection into classes. If, for example, all pupils from the same primary school are placed in the same class at the new secondary school, then the reflection problem is still an issue as is selection into classes. One possible explanation for how classes are formed is that pupils who sign up for similar elective courses (e.g. Latin) are then placed in the same math or Dutch class. Another possibility is that high ability pupils are grouped together. I look at class level performance within each school and find that at 39 of the 57 schools in my sample, the highest performing math class is the same as the highest performing Dutch class. This would imply that high ability pupils are either explicitly grouped together or choose the same electives and therefore end up in the same classes. Of these schools, 31 also have their highest IQ pupils in the best math and Dutch class. Ultimately this type of selection proves much more difficult to address and is something I cannot fully account for without the inclusion of class fixed effects, which means I am cautious in interpreting my results as causal.

The third major challenge to this type of empirical work lies in measurement error. These issues arise because many of the variables used to capture cognitive ability are not very accurate. Grades from the previous school year are often used as a measure of ability, but this ignores the fact that grades actually measure something much more complicated than cognitive ability since they often have a subjective component, such as classroom participation, which is a combination of cognitive and non-cognitive ability. The IRT scores I use as outcome and control measures should suffer less from measurement error since they have been standardised by the LOSO researchers to take into account the difficulty of test items.

In this chapter, I use an IQ test to measure underlying cognitive ability and teacher-reported measures of non-cognitive ability. These non-cognitive measures come from a questionnaire the teacher is required to fill out about each pupil.[31] While there is still some debate as to how accurately IQ measures cognitive ability amongst researchers who focus on latent variable analysis, it is a more accurate measure than previous grades or subject specific test scores. If peer ability is measured with measurement error, this again causes peer effects to be downward biased.

With all of these considerations in mind, I begin the analysis in this chapter with the standard linear-in-means peer effects model:

$$\begin{aligned} Y_{ics,1} &= \beta_1 IQ_{ics,0} + \beta_2 \overline{IQ}_{-ics,0} + \beta_3 NC_{ics,0} + \beta_4 \overline{NC}_{-ics,0} + \beta_5 X_{ics,0} + \beta_6 \overline{X}_{-ics,0} \\ &\quad + \alpha_s + \varepsilon_{icst,1} \end{aligned} \tag{20}$$

[31]The actual measures used in this analysis were developed in conjunction with Dr. Katarzyna Kubacka from the OECD as part of their "Education and Social Progress" project. Further information on these measures may be found in the Appendix.

In this model, $Y_{ics,1}$, is the outcome measure: scores in either Dutch or math at the end of the first year of secondary school. Here $t = 0$ indicates the beginning of the secondary school year and $t = 1$ represents the end of the school year. I include both individual IQ, $IQ_{ics,0}$, measured at the beginning of secondary school, and the class average IQ score, excluding the individual, $\overline{IQ}_{-ics,0}$. The subscript i denotes the individual, while the peer effect terms have the subscript $-i$ to indicate that they are calculated across all individuals except individual i. This means that every individual will have a different peer effect term. The vector $NC_{ics,0}$ includes measures of non-cognitive skills as reported by the class teacher at the beginning of the school year and the vector $\overline{NC}_{-ics,0}$ includes the averages of these measures for all the peers in a class. Similarly, the vector $X_{ics,0}$ includes measures of background characteristics as well as previous math or Dutch performance and the vector $\overline{X}_{-ics,0}$ includes the averages of these measures for all the peers in a class, excluding individual, i. I include school fixed effects, α_s, in order to account for unobserved heterogeneity at the school level.

There is a large literature on which variables best represent socio-economic status (SES) and how these are then related to educational outcomes. Sirin (2005) conducts a meta-analysis of all papers that have been published between 1990-2000 to examine the strength of various measures in explaining educational achievement. He acknowledges the "tripartite nature of SES that incorporates parental income, parental education, and parental occupation as the three main indicators of SES" (Sirin, 2005), which is how I determine which variables to include in my vector $X_{ics,0}$. I include a binary variable for whether or not the father of the pupil attended tertiary education and information on family income. In this vector I also include pupil specific characteristics such as gender.

I will first estimate a series of basic linear-in-means models with school fixed effects, look at one aspect of heterogeneity between pupils, and then use quantile regression to look at the linear-in-means model at other quantiles of the ability distribution. Following this analysis, I will explore non-linear peer effects models. The non-linear-in-means peer effects models I estimate will allow for variance in ability to enter into the model and also allow for peers in different portions of the ability distribution to have different effects on the individual. The variance model I estimate takes the following form:

$$\begin{aligned} Y_{ics,1} \quad = \quad & \beta_1 IQ_{ics,0} + \beta_2 \overline{IQ}_{-ics,0} + \beta_3 Var(IQ)_{-ics,0} + \beta_4 NC_{ics,0} + \beta_5 \overline{NC}_{-ics,0} \\ & + \beta_6 Var(NC)_{-ics,0} + \beta_7 X_{ics,0} + \beta_8 \overline{X}_{-ics,0} + \alpha_s + \epsilon_{icst,1} \end{aligned} \tag{21}$$

The inclusion of both the mean peer effect term and variance peer effect term for the measures of IQ, non-cognitive ability, and previous math or Dutch performance will allow me to disentangle the effect of the tightness of the ability distribution from the mean on outcomes. Policies of streaming work on the assumption that a class with a lower variance, more similar pupils, positively affects a pupil's outcome. Including the variance terms will allow me to test this.

There are also models of peer effects that focus on the impact of having very high ability or very low ability pupils in a classroom. The "shining light" model hypothesises that having very good pupils in the classroom serves as an example for the others and thereby pulls them all up, while the "bad apple" model hypothesises that having very poor performing or ill-behaved pupils disturbs learning in such a way to drag her peers down (Sacerdote, 2001). In order to test the validity of these two hypotheses, I estimate a model that includes the percent of pupils in each class in the bottom and top quintiles of the entire cohort's ability distribution. I include these quintile variables for both cognitive and non-cognitive measures in the following model:

$$\begin{aligned} Y_{ics,1} &= \beta_1 IQ_{ics,0} + \beta_2 \overline{IQ}_{-ics,0} + \beta_3 PercentTopQuintile(IQ)_{-ics,0} \\ &+ \beta_4 PercentBottomQuintile(IQ)_{-ics,0} + \beta_5 NC_{ics,0} + \beta_6 \overline{NC}_{-ics,0} \\ &+ \beta_7 PercentTopQuintile(NC)_{-ics,0} + \beta_8 PercentBottomQuintile(NC)_{-ics,0} \\ &+ \beta_9 X_{ics,0} + \beta_{10} \overline{X}_{-ics,0} + \alpha_s + \xi_{icst,1} \end{aligned} \tag{22}$$

I cluster the standard errors in all models at the classroom level in order to obtain standard errors robust to potential intra-cluster correlation. In the following Results section, I present the results from the aforementioned models in the same order discussed here.

4.7 Results

In this section of the chapter, I present the results from the models discussed in the Empirical Strategy section. Because of the issue of school choice, I always include school fixed effects in all regressions. It should be noted that by including school fixed effects, my identification is being driven by the differences across classes. This is why having two or more classes at a school is fundamental to this identification strategy. The descriptive statistics show that there are on average six classes per school and at least two at every school, which should allow for enough heterogeneity across classes in order to obtain identification.

Following the tables for the linear-in-means models, I include a table of heterogeneous peer effects for pupils of varying IQ. Then I present the results of using quantile regression to estimate the linear-in-means model at the 10th, 50th, and 90th quantiles of the ability distribution. Following this, I present the results from the non-linear-in-means models of peer effects. This includes a table presenting the variance model discussed in the Empirical Strategy section as well as the non-linear peer effects model that uses quintiles of peer ability in a classroom.

4.7.1 Linear-in-means Models

In each table presented in this section for the linear-in-means model results there are five columns. Each of these five columns represents a slightly different specification of a linear model[32] and the order is constant across tables for the sake of clarity. I group the peer effect terms (denoted by the prefix “mean”) together and the individual level controls together in each table. In Column (1) of each table I run a basic OLS regression that represents a “value added model.” This is a standard model within the literature on academic performance. Here past math or Dutch achievement, IQ, gender, SES, and class size are regressed on math or Dutch achievement at the end of the school year. This value added specification gives us an idea of what the baseline model without any inclusion of peer effects looks like and serves as a reference point for comparison.

In Column (2) of each table I introduce individual level non-cognitive measures to the baseline value added model. This extends the traditional value added model because generally such measures of personality traits are not available. I include these measures so that we may differentiate the effect of individual level non-cognitive traits from the effect of non-cognitive peer effects once they are added to the baseline.

Column (3) presents the results of the traditional linear-in-means peer effects model common in the literature (Sacerdote, 2001). This model includes all of the variables from the baseline in Column (1), but also includes the peer effect averages for IQ, previous math or Dutch performance, income, and proportion of males in the classroom. I do not include any non-cognitive measures in this specification because I want to be able to compare my model with non-cognitive peer effects to the traditional peer effects model. In Column (4) I add individual level non-cognitive measures to the traditional linear-in-means peer effects model in order to differentiate between individual and peer level non-cognitive traits. I am also interested to see if the inclusion of individual level non-cognitive traits changes the results of the traditional peer effects model.

Finally, in Column (5) of each table I present the results from the full model. This includes all of the baseline variables and traditional peer effects terms plus individual non-cognitive measures and the average of these measures across peers. This is the model of interest as it fully extends the baseline to include both cognitive and non-cognitive peer and individual effects. Across all specifications for math and Dutch, the coefficients on my variables of interest stay constant as do the other parameters, which gives me confidence in the validity of the approach I am taking.

The results of the linear-in-means model of the math scores show that the peer effects terms vary in their significance and sign. The only non-cognitive peer effects to have a statistically significant impact on individual level learning outcomes are conscientiousness and extraversion. The peer effect

[32] This logic does not apply to the tables containing the results for the heterogeneous peer effects model, quantile regression, variance model, and the non-linear peer effects model.

Table 50: Linear-in-Means Model: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score	(5) math score
mean_math_irtscore_t0_m			0.1082	0.0056	-0.0292
			(0.0804)	(0.0779)	(0.0784)
mean_iq_m			0.0007	0.0053	-0.0019
			(0.0064)	(0.0065)	(0.0067)
mean_CON_m					0.0733***
					(0.0230)
mean_AGREE_m					0.0205
					(0.0243)
mean_EXTRA_m					-0.0630**
					(0.0270)
mean_male_m			-0.2171	-0.0926	-0.0151
			(0.1637)	(0.1716)	(0.1668)
mean_father_teredu_m			0.7542***	0.5131***	0.4370**
			(0.1810)	(0.1971)	(0.2013)
math_irtscore_t0	0.3885***	0.3572***	0.3582***	0.3442***	0.3426***
	(0.0277)	(0.0285)	(0.0268)	(0.0280)	(0.0282)
getlov_iq	0.0291***	0.0231***	0.0259***	0.0217***	0.0216***
	(0.0020)	(0.0024)	(0.0020)	(0.0024)	(0.0025)
CON		0.0691***		0.0609***	0.0592***
		(0.0089)		(0.0093)	(0.0091)
AGREE		0.0043		0.0045	0.0057
		(0.0065)		(0.0065)	(0.0065)
EXTRA		-0.0109		-0.0104	-0.0103
		(0.0071)		(0.0071)	(0.0071)
male	-0.1107***	-0.0528	-0.0805**	-0.0439	-0.0411
	(0.0391)	(0.0485)	(0.0385)	(0.0483)	(0.0483)
father_teredu	0.0758**	0.0255	0.0558*	0.0152	0.0161
	(0.0314)	(0.0352)	(0.0304)	(0.0345)	(0.0346)
math_class_size	-0.0143**	-0.0184***	-0.0235***	-0.0222***	-0.0217***
	(0.0065)	(0.0068)	(0.0072)	(0.0078)	(0.0077)
Constant	-3.1678***	-3.1772***	-2.6590***	-3.4189***	-3.0885***
	(0.2324)	(0.2512)	(0.7236)	(0.7331)	(0.7389)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	4,894	3,256	4,894	3,256	3,255
R-squared	0.5191	0.5489	0.5267	0.5526	0.5552

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income not presented in this table

term for conscientiousness positively and significantly impacts individual level learning outcomes. This effect is 0.073 points on the math score in the full model, Column (5) of Table 50. This is equivalent to a one standard deviation increase in peer level conscientiousness being associated with a 0.14 standard deviation increase in math scores, which reinforces the idea that having a positive learning environment is related to better math outcomes.

The peer effect for extraversion is statistically significant and negative in Table 50 with a magnitude of -0.063, which is the same as a one standard deviation increase in average peer extraversion being associated with a -0.065 standard deviation decrease in math scores. If we interpret the peer effect on extraversion to be highly correlated with a disruptive learning environment due to the talkative nature of the pupils, then peer level extraversion is negatively related to individual learning outcomes. The peer effect of agreeableness is not statistically significant; however, it is positive, which is in accordance with our beliefs about agreeable peers fostering a positive learning environment. These results indicate that the learning environment, as created by one's peers, is related to learning outcomes. This result is similar to Neidell and Waldfogel's finding that peer externalising problems, which they describe as "most likely capture[ing] classroom disturbance", negatively affect learning outcomes.

The peer effect term for IQ is consistently small and not statistically significant. This means that high average peer IQ does not improve learning outcomes. This finding contradicts much of the existing literature previously discussed, which finds a positive impact of IQ on learning outcomes. Most other peer effects studies do not control for conscientiousness, which is positively correlated with IQ, so it is possible that the significant findings on IQ in other studies are actually proxying for conscientiousness.

The peer effect term for having peers with educated fathers is also statistically significant and positive across specifications, unlike the peer effect terms for income (not reported in this table), which are never statistically significant in any specification. Having peers from more educated families is positively related to math outcomes.

These results also show the importance of individual level non-cognitive measures on math scores. Across all specifications, conscientiousness at the individual level is statistically significant and positive. In the literature, conscientiousness has been found to be positively correlated with educational and labour market outcomes (Borghans et al., 2008). Poropat (2009) conducts a meta-analysis of studies using the Big Five model of personality and academic outcomes and finds that the correlation between conscientiousness and academic performance is independent of IQ and that conscientiousness predicts as much of tertiary education performance as IQ once past schooling performance is included in the model.

The introduction of the individual level non-cognitive measures in Column (2) of Table 50 is interesting because the general results from the baseline do not change. The effect of individual level IQ is approximately 0.02 points, and also remains statistically significant and roughly the same size throughout all specifications. The standardised effect is a one standard deviation increase in individual IQ being associated with a 0.24 standard deviation increase in math scores. Compared to the non-cognitive peer effects, individual IQ has a much stronger relationship with math scores, which is to be expected. The fact that the coefficient on individual IQ does not change across specifications means that IQ is not proxying for non-cognitive ability and that these measures actually are measuring something independently important. Individual past performance in math positively predicts math performance with a coefficient of 0.35 points, which also remains relatively constant in terms of size and statistical significance across specifications. This is the same as a one standard deviation increase in past math performance being associated with a 0.30 standard deviation increase in math scores.

The role of gender in the math score regressions reveals an interesting result also shown in the literature on non-cognitive traits. Overall, boys perform worse in math than girls; however, this result only holds in models without non-cognitive measures. As discussed in the Literature Review, gender may often proxy for non-cognitive traits (Bertrand and Pan, 2011), which seems to be the case here. Once non-cognitive traits are included, the negative relationship between being a boy and math performance disappears.

This negative performance of boys in the standard models is still surprising in the Belgian context where in 2009, boys outperformed girls on the PISA mathematics test by more than 20 points, which is the same as one third of a proficiency level (OECD, 2011a). The gap still existed on the 2012 PISA, where Flemish boys, a subset of all Belgian pupils participating in PISA, outperformed Flemish girls by 12 points (OECD, 2013). The PISA tests come nearly 20 years after the LOSO project and actually test a different age group (15 year olds take the PISA and the children in the LOSO project are 12-13 years old), which could explain these differing results.

One variable that negatively predicts math scores across all models is class size. In all regressions, the variable for math class size is statistically significant and negative, indicating that a one standard deviation increase in class size is associated with a 0.07 standard deviation decrease in math performance. Regardless of the composition or characteristics of the class, having a larger class is not a good thing. Again, this reinforces the idea that the environment of the classroom, in this case defined by size, matters quite a lot for math achievement.

The results for the regressions using Dutch scores have some key differences and some similarities to the math results. At the individual level and the peer level, the non-cognitive traits that matter differ from math to Dutch. The results of the full model in Column (5) of Table 51 show that peer level conscientiousness is also positively related to individual performance, as was the case with math. A one standard deviation increase in average peer conscientiousness is associated with a 0.129 standard deviation increase in Dutch scores. This supports the idea that having diligent

Table 51: Linear-in-Means Model: Dutch Scores

VARIABLES	(1) Dutch score	(2) Dutch score	(3) Dutch score	(4) Dutch score	(5) Dutch score
mean_dutch_irtscore_t0_d			0.0916	0.0369	-0.0252
			(0.0900)	(0.0823)	(0.0823)
mean_iq_d			-0.0004	0.0014	-0.0035
			(0.0058)	(0.0064)	(0.0063)
mean_CON_d					0.0577***
					(0.0184)
mean_AGREE_d					-0.0075
					(0.0174)
mean_EXTRA_d					0.0009
					(0.0198)
mean_male_d			-0.1278	-0.0827	-0.0321
			(0.1372)	(0.1445)	(0.1454)
mean_father_teredu_d			0.6219***	0.4023**	0.2662
			(0.1495)	(0.1675)	(0.1663)
dutch_irtscore_t0	0.4664***	0.4265***	0.4378***	0.4139***	0.4170***
	(0.0360)	(0.0407)	(0.0342)	(0.0405)	(0.0395)
getlov_iq	0.0221***	0.0179***	0.0202***	0.0171***	0.0169***
	(0.0016)	(0.0017)	(0.0015)	(0.0016)	(0.0016)
CON		0.0551***		0.0497***	0.0469***
		(0.0063)		(0.0064)	(0.0059)
AGREE		-0.0005		-0.0005	0.0005
		(0.0040)		(0.0040)	(0.0040)
EXTRA		-0.0121**		-0.0118**	-0.0118**
		(0.0049)		(0.0049)	(0.0047)
male	-0.2604***	-0.2416***	-0.2471***	-0.2374***	-0.2328***
	(0.0307)	(0.0326)	(0.0296)	(0.0331)	(0.0328)
father_teredu	0.0843***	0.0652**	0.0708***	0.0591**	0.0548*
	(0.0258)	(0.0301)	(0.0251)	(0.0295)	(0.0292)
dutch_class_size	-0.0131**	-0.0178***	-0.0188***	-0.0210***	-0.0204***
	(0.0053)	(0.0053)	(0.0052)	(0.0058)	(0.0057)
Constant	-1.8537***	-1.9207***	-1.4635**	-1.8375***	-1.8762***
	(0.2334)	(0.2603)	(0.6553)	(0.6855)	(0.6785)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	4,907	3,269	4,906	3,269	3,268
R-squared	0.7015	0.7049	0.7078	0.7079	0.7100

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income not presented in this table

peers benefits an individual's learning outcomes. None of the other non-cognitive peer effects have a statistically significant relationship with individual Dutch outcomes.

The peer effect term for having peers whose fathers attended tertiary education is large, positive, and statistically significant in Columns (3) and (4) of Table 51. Once I add in peer level non-cognitive measures, this relationship vanishes. This indicates that much of the benefit individuals get from having peers from better educated families may actually be transmitted through their non-cognitive traits.

Unlike in math, the peer effect term for extraversion does not have a statistically significant association with Dutch performance. The potentially disruptive nature of more extroverted peers is not related to Dutch learning outcomes.

At the individual level, there is a statistically significant and positive relationship between conscientiousness and Dutch performance of approximately 0.05 points; this is the same as a one standard deviation increase in individual conscientiousness being associated with a 0.144 standard deviation increase in Dutch scores. Again, this is in line with the literature on conscientiousness being highly related to labour market and academic success (Borghans et al., 2008). The results for individual level extraversion are also similar to math: in the Dutch specifications, there is a statistically significant and negative relationship between being more extroverted and Dutch performance. In the Dutch regressions, agreeableness does not predict performance in a statistically significant way.

A key difference between the math and Dutch results is the role of gender. In all of the Dutch specifications across tables, there is a statistically significant and negative relationship between being a boy of approximately -0.24 points on Dutch performance. This does not change once non-cognitive measures are introduced. The significance of the gender effect here is actually in line with the 2012 PISA results, in which girls in Belgium outperformed boys in reading by an average of 32 points (OECD, 2013). Girls in Belgium have consistently performed better on reading in PISA than boys, but worse than boys on math (OECD, 2013). As was the case in the math regressions, there is no gender peer effect; having more boys in a classroom does not affect individual level learning outcomes. This is different from the gender peer effect found in Fraine et al. (2003) and Opdenakker et al. (2002) using the LOSO data; however, since they do not use a linear-in-means model and do not exclude the individual from the peer level measure, it is difficult to compare our findings.

4.7.2 Heterogeneous Treatment Effects

I explore the possibility of heterogeneous treatment effects for pupils based on ability by including an interaction term of a dummy for whether or not the individual's IQ is above the median across the whole sample with the peer effect term for IQ in her class. Later on I will also estimate peer effects at different quantiles of the ability distribution, but this method allows me to estimate heterogeneous treatment effects at the mean. These results are shown in Tables 52 and 53 for math and Dutch respectively. Unlike the previous tables, Tables 52 and 53 only include three columns each. Column (1) is the traditional peer effects model, Column (2) is the traditional peer effects model plus individual level non-cognitive measures, and Column (3) is the complete model with both cognitive and non-cognitive individual and peer effect terms. In each of these columns, I include my interaction term for individual level IQ being above the median and the peer effect term for IQ.[33]

[33] I do not include any of the models without peer effects terms in these tables since I am interested in the interaction of the peer effect term for IQ and individual IQ.

Table 52: Linear-in-Means Model with Heterogeneous Effects: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score
mean_math_irtscore_t0_m	0.1166	0.0098	-0.0242
	(0.0783)	(0.0772)	(0.0781)
mean_iq_m	0.0002	0.0050	-0.0019
	(0.0062)	(0.0065)	(0.0067)
above_median_iq*mean_iq_m	0.0017***	0.0010*	0.0009*
	(0.0005)	(0.0005)	(0.0005)
mean_CON_m			0.0709***
			(0.0228)
mean_AGREE_m			0.0201
			(0.0242)
mean_EXTRA_m			-0.0622**
			(0.0269)
mean_male_m	-0.2044	-0.0800	-0.0071
	(0.1591)	(0.1694)	(0.1652)
mean_father_teredu_m	0.6860***	0.4808**	0.4130**
	(0.1758)	(0.1967)	(0.2014)
math_irtscore_t0	0.3583***	0.3452***	0.3435***
	(0.0266)	(0.0280)	(0.0282)
getlov_iq	0.0212***	0.0188***	0.0191***
	(0.0025)	(0.0032)	(0.0032)
CON		0.0594***	0.0579***
		(0.0093)	(0.0092)
AGREE		0.0049	0.0060
		(0.0065)	(0.0065)
EXTRA		-0.0096	-0.0096
		(0.0072)	(0.0071)
male	-0.0809**	-0.0449	-0.0420
	(0.0384)	(0.0483)	(0.0483)
father_teredu	0.0513*	0.0156	0.0165
	(0.0305)	(0.0348)	(0.0348)
math_class_size	-0.0221***	-0.0214***	-0.0211***
	(0.0070)	(0.0077)	(0.0076)
Constant	-2.2330***	-3.1633***	-2.8715***
	(0.7317)	(0.7591)	(0.7620)
School FE	Yes	Yes	Yes
Observations	4,894	3,256	3,255
R-squared	0.5286	0.5533	0.5557

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income not presented in this table

Table 53: Linear-in-Means Model with Heterogeneous Effects: Dutch Scores

VARIABLES	(1) Dutch score	(2) Dutch score	(3) Dutch score
mean_dutch_irtscore_t0_d	0.0990	0.0388	-0.0204
	(0.0867)	(0.0806)	(0.0810)
mean_iq_d	-0.0010	0.0011	-0.0035
	(0.0057)	(0.0063)	(0.0063)
above_median_iq*mean_iq_d	0.0016***	0.0011***	0.0009***
	(0.0003)	(0.0003)	(0.0003)
mean_CON_d			0.0551***
			(0.0178)
mean_AGREE_d			-0.0081
			(0.0173)
mean_EXTRA_d			0.0013
			(0.0196)
mean_male_d	-0.1215	-0.0747	-0.0276
	(0.1335)	(0.1421)	(0.1435)
mean_father_teredu_d	0.5617***	0.3692**	0.2419
	(0.1419)	(0.1633)	(0.1638)
dutch_irtscore_t0	0.4413***	0.4174***	0.4201***
	(0.0335)	(0.0403)	(0.0395)
getlov_iq	0.0157***	0.0141***	0.0142***
	(0.0016)	(0.0019)	(0.0019)
CON		0.0480***	0.0455***
		(0.0063)	(0.0060)
AGREE		-0.0001	0.0008
		(0.0040)	(0.0040)
EXTRA		-0.0109**	-0.0110**
		(0.0049)	(0.0047)
male	-0.2465***	-0.2370***	-0.2326***
	(0.0296)	(0.0331)	(0.0328)
father_teredu	0.0673***	0.0603**	0.0560*
	(0.0251)	(0.0295)	(0.0293)
dutch_class_size	-0.0172***	-0.0200***	-0.0196***
	(0.0050)	(0.0057)	(0.0056)
Constant	-1.0647*	-1.5776**	-1.6386**
	(0.6374)	(0.6882)	(0.6845)
School FE	Yes	Yes	Yes
Observations	4,906	3,269	3,268
R-squared	0.7098	0.7088	0.7107

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income not presented in this table

Table 52 shows that the interaction term for being above the median IQ with peer IQ is positive and statistically significant. In the full model in Column (3), this means that pupils with above median IQ are hurt less by having higher performing peers than their counterparts. In the models without the non-cognitive peer effects, they benefit more.

Table 53 similarly shows a less negative impact of having higher ability peers for pupils above the median IQ on their Dutch performance in Column (3) and positive effects of higher ability peers in Columns (1) and (2).

4.7.3 Quantile Regression

In this section of the chapter, I use quantile regression to estimate the same linear-in-means model specified in Column (5) of the linear-in-means tables, but for different conditional quantiles. As Koenker and Hallock (2001) point out, quantile regression is an estimation technique used for conducting inference on conditional quantile functions. This is different from linear regression, which allows us to estimate models of the conditional mean. The primary difference between linear regression and quantile regression is that in quantile regression we solve the minimisation problem of some weighted sum of absolute residuals as opposed to the sum of squared residuals (Koenker and Hallock, 2001). This allows us to estimate the impact of independent variables at certain points of the conditional distribution of the outcome variable. In a context such as this one, it seems prudent to look beyond the mean at various quantiles of the distribution to see how peer effects affect low and high ability pupils differently.

In Table 54, I present the results of this quantile regression estimation on the math scores and in Table 55, I present the Dutch score results. As a point of reference, in the first column of each table, I include the full linear-in-means peer effects model already presented in Column (5) of Tables 50 and 51. This allows us to compare the mean estimate with those estimates at various quantiles. The next three columns present the quantile regression results for the 10th, 50th, and 90th quantiles respectively.

The results in Table 54, indicate that some of the peer effects affect pupils at different points of the conditional outcome distribution differently. I find that that peer effect on conscientiousness, which is positive and statistically significant in the linear regression is only positive and statistically significant for the 10th and 50th quantiles. It also decreases in size as we move along the conditional outcome distribution. The fact that peer level conscientiousness matters more for those at the bottom of the ability distribution might mean that conscientiousness serves as a coping mechanism to compensate for their lower cognitive ability. These pupils benefit more than their peers from being surrounded by conscientious people because it creates a positive learning environment.

The peer effect on extraversion, which is negative and statistically significant in the linear regression is only statistically significant for pupils in the 90th quantile. This indicates a disruptive learning environment is worse for high ability pupils. These results in Table 54 also show that the negative

Table 54: Quantile Regression: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score
	Linear regression		Quantile Regression	
		10th	50th	90th
mean_math_irtscore_t0_m	-0.0292	-0.1357	0.0668	0.0136
	(0.0784)	(0.2096)	(0.0725)	(0.0623)
mean_iq_m	-0.0019	-0.0121	-0.0008	0.0063
	(0.0067)	(0.0141)	(0.0057)	(0.0058)
mean_CON_m	0.0733***	0.0907*	0.0313*	0.0235
	(0.0230)	(0.0532)	(0.0181)	(0.0176)
mean_AGREE_m	0.0205	0.0217	0.0258	0.0322*
	(0.0243)	(0.0596)	(0.0196)	(0.0172)
mean_EXTRA_m	-0.0630**	-0.0596	-0.0257	-0.0377**
	(0.0270)	(0.0608)	(0.0240)	(0.0188)
mean_male_m	-0.0151	-0.1242	-0.0630	0.0925
	(0.1668)	(0.2575)	(0.1084)	(0.1307)
mean_father_teredu_m	0.4370**	0.4271	0.2239	0.0082
	(0.2013)	(0.4647)	(0.1468)	(0.1944)
math_irtscore_t0	0.3426***	0.4868***	0.3705***	0.2956***
	(0.0282)	(0.0572)	(0.0228)	(0.0209)
getlov_iq	0.0216***	0.0263***	0.0145***	0.0149***
	(0.0025)	(0.0039)	(0.0017)	(0.0015)
CON	0.0592***	0.0788***	0.0522***	0.0334***
	(0.0091)	(0.0190)	(0.0077)	(0.0075)
AGREE	0.0057	0.0281*	-0.0003	0.0071
	(0.0065)	(0.0154)	(0.0055)	(0.0052)
EXTRA	-0.0103	-0.0207	-0.0110**	-0.0087
	(0.0071)	(0.0190)	(0.0055)	(0.0055)
male	-0.0411	-0.0074	-0.0065	-0.0461
	(0.0483)	(0.0960)	(0.0396)	(0.0349)
father_teredu	0.0161	0.0757	-0.0044	-0.0061
	(0.0346)	(0.0671)	(0.0354)	(0.0260)
math_class_size	-0.0217***	-0.0255*	-0.0079	-0.0042
	(0.0077)	(0.0134)	(0.0056)	(0.0059)
Constant	-3.0885***	-4.1117***	-2.3695***	-2.4431***
	(0.7389)	(1.4920)	(0.5772)	(0.6007)
School FE	Yes	Yes	Yes	Yes
Observations	3,255	3,255	3,255	3,255
R-squared	0.5552	0.5214	0.5415	0.5175

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income not presented in this table

Table 55: Quantile Regression: Dutch Scores

VARIABLES	(1) Dutch score	(2) Dutch score	(3) Dutch score	(4) Dutch score
	Linear regression		Quantile Regression	
		10th	50th	90th
mean_dutch_irtscore_t0_d	-0.0252	0.0797	0.0240	-0.0340
	(0.0823)	(0.1858)	(0.0828)	(0.1903)
mean_iq_d	-0.0035	0.0011	-0.0053	-0.0027
	(0.0063)	(0.0100)	(0.0056)	(0.0076)
mean_CON_d	0.0577***	0.0337	0.0304	0.0726
	(0.0184)	(0.0322)	(0.0236)	(0.0498)
mean_AGREE_d	-0.0075	-0.0009	-0.0139	-0.0081
	(0.0174)	(0.0343)	(0.0188)	(0.0294)
mean_EXTRA_d	0.0009	-0.0233	0.0124	-0.0079
	(0.0198)	(0.0396)	(0.0211)	(0.0279)
mean_male_d	-0.0321	-0.0765	-0.1104	-0.1123
	(0.1454)	(0.2151)	(0.1171)	(0.1477)
mean_father_teredu_d	0.2662	-0.1258	0.4024**	0.1860
	(0.1663)	(0.2767)	(0.1964)	(0.1706)
dutch_irtscore_t0	0.4170***	0.5028***	0.4443***	0.3840***
	(0.0395)	(0.0442)	(0.0230)	(0.0412)
getlov_iq	0.0169***	0.0192***	0.0165***	0.0149***
	(0.0016)	(0.0023)	(0.0014)	(0.0022)
CON	0.0469***	0.0359***	0.0469***	0.0497***
	(0.0059)	(0.0105)	(0.0066)	(0.0091)
AGREE	0.0005	0.0066	0.0019	-0.0045
	(0.0040)	(0.0097)	(0.0047)	(0.0080)
EXTRA	-0.0118**	-0.0098	-0.0121***	-0.0022
	(0.0047)	(0.0098)	(0.0046)	(0.0078)
male	-0.2328***	-0.2895***	-0.1640***	-0.1213***
	(0.0328)	(0.0524)	(0.0340)	(0.0385)
father_teredu	0.0548*	0.0266	0.1067***	0.0140
	(0.0292)	(0.0633)	(0.0252)	(0.0439)
dutch_class_size	-0.0204***	-0.0086	-0.0162***	-0.0189***
	(0.0057)	(0.0086)	(0.0060)	(0.0070)
Constant	-1.8762***	-2.8255**	-1.4573**	-1.4539
	(0.6785)	(1.0977)	(0.5910)	(0.9183)
School FE	Yes	Yes	Yes	Yes
Observations	3,268	3,268	3,268	3,268
R-squared	0.7100	0.6855	0.7043	0.6842

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income not presented in this table

relationship with class size only applies to the 10th quantile, indicating that low ability pupils suffer more from being in a large class. The fact that these two variables affect different quantiles differently shows that class size and peer level extraversion are not the same thing. The negative aspects of being in a large class might arise from being neglected from the teacher, not by how disruptive it might be.

At the individual level, the measure for agreeableness is only statistically significant and positive for the 10th quantile. Conscientiousness on the other hand is statistically significant and positive across the distribution, but decreasing in magnitude as the quantiles increase. Again, this may point to using conscientiousness as a coping strategy for low cognitive ability.

The quantile regression results for the Dutch scores shown in Table 55 have some key differences. None of the non-cognitive peer effect variables are statistically significant for different quantiles, unlike in the linear regression. The positive and statistically significant peer effect of conscientiousness found in the linear regression is not statistically significant across any of the quantiles. The peer effect term of having an educated father is positive and statistically significant at the 50th quantile, which indicates that median ability pupils benefit from being surrounded by high socio-economic status peers.

At the individual level, the results on the non-cognitive measures are similar to the math regressions. Conscientiousness positively predicts performance for pupils across the distribution and has the strongest association for pupils at the 90th quantile. This does not support the idea of using conscientiousness as a coping strategy for lower cognitive ability, but points to more of a synergy between the two. The individual level effect of extraversion is only negative for pupils at the 50th quantile of the distribution. Unlike math, agreeableness at the individual level is not statistically significant for any quantile. In the Dutch regressions, the boys perform worse at every quantile, with the relationship largest at the 10th quantile and decreasing in magnitude as we move up the quantiles.

The results of this quantile regression show some interesting results not captured by estimating the linear-in-means model through linear regression. Pupils at different quantiles of the ability distribution react differently to their peers.

4.7.4 Variance Model and Non-linear Peer Effects

I now turn my attention to the issue of non-linear peer effects. Until now, all of the models I have estimated have been linear-in-means models, which force the peer effect term to enter the model in a linear way for all pupils. They also restrict the interpretation of peer effects to the average of peer ability. These restrictions might not be realistic if we believe that other characteristics of the ability distribution of peers within a class might matter. Perhaps having a low variance in ability is better for learning outcomes because teachers are able to teach to one level? Alternatively, if we believe the shining light or bad apple model, then having a higher percentage of very high and very low-performing pupils might also affect learning outcomes in a way the traditional linear-in-means model is unable to measure.

In order to examine these alternative models of peer effects, I first estimate the same linear-in-means model as before, but include a variance term for each of the cognitive and non-cognitive peer effects terms. This is the model presented in Equation (21). This variance term measures the variance of all peers' ability, excluding the individual pupil.

The results of this estimation on the math scores do not reveal any major differences. None of the variance terms in Column (1) of Table 56 are statistically significant; however, the mean peer effect terms for conscientiousness and extraversion still have the same signs they had in the linear-in-means models presented in Table 50. The average of peer level conscientiousness has

Table 56: Variance Model

VARIABLES	(1) math score	VARIABLES	(2) Dutch score
math_irtscore_t0	0.3459***	dutch_irtscore_t0	0.4201***
	(0.0280)		(0.0395)
mean_math_irtscore_t0_m	0.0374	mean_dutch_irtscore_t0_d	0.0067
	(0.0945)		(0.0884)
var_math_irtscore_t0_m	0.0423	var_dutch_irtscore_t0_d	0.0198
	(0.0384)		(0.0359)
getlov_iq	0.0209***	getlov_iq	0.0164***
	(0.0024)		(0.0015)
mean_iq_m	-0.0063	mean_iq_d	-0.0059
	(0.0074)		(0.0058)
var_iq_m	-0.0007	var_iq_d	-0.0001
	(0.0006)		(0.0004)
CON	0.0615***	CON	0.0481***
	(0.0090)		(0.0059)
mean_CON_m	0.0682***	mean_CON_d	0.0508**
	(0.0245)		(0.0217)
var_CON_m	0.0001	var_CON_d	-0.0102*
	(0.0065)		(0.0054)
AGREE	0.0056	AGREE	0.0010
	(0.0065)		(0.0040)
mean_AGREE_m	0.0234	mean_AGREE_d	-0.0004
	(0.0317)		(0.0206)
var_AGREE_m	-0.0005	var_AGREE_d	0.0035
	(0.0067)		(0.0056)
EXTRA	-0.0100	EXTRA	-0.0110**
	(0.0071)		(0.0048)
mean_EXTRA_m	-0.0595**	mean_EXTRA_d	0.0148
	(0.0275)		(0.0221)
var_EXTRA_m	-0.0047	var_EXTRA_d	0.0072*
	(0.0069)		(0.0042)
father_teredu	0.0228	father_teredu	0.0593**
	(0.0336)		(0.0284)
mean_father_teredu_m	0.4273**	mean_father_teredu_d	0.1700
	(0.2066)		(0.1691)
math_class_size	-0.0213***	dutch_class_size	-0.0187***
	(0.0078)		(0.0057)
Constant	-2.5569***	Constant	-1.7534***
	(0.7547)		(0.6192)
School FE	Yes	School FE	Yes
Observations	3,244	Observations	3,256
R-squared	0.5573	R-squared	0.7130

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income and gender not presented in this table

a statistically significant and positive relationship with math scores and the average peer level measure of extraversion has a negative and statistically significant relationship with math scores.

The results of the variance model for Dutch in Table 56 are different. Here the peer effect variance term on conscientiousness is significant at the 10 percent significance level and negative. A one standard deviation increase in the variance of peer conscientiousness variable is associated with a -0.032 standard deviation decrease in Dutch scores. The mean peer effect term of conscientiousness is still positive and statistically significant, with a one standard deviation increase in average peer conscientiousness being associated with a 0.113 standard deviation increase in Dutch scores. This indicates that having more conscientious peers on average is beneficial, but that having a greater dispersion of conscientiousness in a classroom hurts individual learning outcomes. This result supports the idea of implementing interventions to improve pupil conscientiousness. In both the Dutch and the math regressions, having a higher variance of cognitive ability, both in terms of past performance and IQ, does not affect learning outcomes in a statistically significant way.

These results do not rule out the possibility of other non-linear peer effects. I now turn my attention to the effect of having more very high or very low ability peers in a classroom. In order to test this, I take the distribution of all pupils in the data set for a variety of measures and divide it into five quintiles. Using these quintiles, I then calculate the percentage of classmates a given pupil has from each quintile. This excludes the pupil's own place in the distribution from these measures and gives me a relative measure of how many peers fall into each quintile that I can compare across pupils. In Column (1) of each of the quintile model tables I present the full linear-in-means model, in Column (2) I include the top and bottom quintile measures, but do not control for the average of each variable, and in Column (3) I include the top and bottom quintile measures as well as the average peer effect term.

The results in Column (2) of Table 57 indicate that there is a relationship between having more peers in the top quintile of agreeableness and math scores. A one standard deviation increase in the percentage of top quintile agreeable peers is associated with a 0.043 standard deviation increase in math scores. Having more very agreeable peers, which may benefit the learning environment, is related to higher individual outcomes. There is also a statistically significant and positive relationship between having more peers in the bottom quintile of extraversion in Column (2) and math scores, meaning that more introverted peers positively predict individual math performance. This is equivalent to a one standard deviation increase in percentage of bottom quintile extroverted peers being associated with a 0.047 standard deviation increase in math scores. These associations are no longer statistically significant when I include the average peer effect terms in Column (3).

The results for the Dutch scores in Table 58 show some interesting non-linear relationships. Although there was no effect of average peer Dutch performance from the beginning of the school year, there is a positive and statistically significant relationship between having more very high Dutch performers on individual scores. This coefficient on the top quintile of Dutch performance is positive and statistically significant in Columns (2) and (3), lending support to the shining light model of peer effects for previous ability. A one standard deviation increase in percentage of top quintile performing peers in Dutch is associated with a 0.084 standard deviation increase in individual Dutch scores in Column (2) and a 0.090 standard deviation increase in individual Dutch scores in Column (3).

In Column (3), the top quintile of IQ variable is also positive and statistically significant, further supporting the shining light model. Here a one standard deviation increase in percentage of peers in the top quintile of IQ is associated with a 0.077 standard deviation increase in Dutch scores. Burke and Sass (2013) use a similar identification strategy, employing school fixed effects, and also find no statistically significant impact of their peer effect term of interest, previous peer academic performance, in the linear-in-means context, but find non-linear effects.

Turning to the non-cognitive peer effects, these results show that having more peers in the bottom quintile of conscientiousness is negatively related to individual Dutch scores in Columns (2) of Table 58. This effect of having bad apples disappears once I control for average peer conscientiousness in the class. There is also a statistically significant and negative relationship in Column (3) of having more peers from the top quintile of the conscientiousness distribution with Dutch scores. This result complements the finding from the variance model in Table 56, which showed that having a greater variance of conscientiousness in a classroom hurts individual outcomes.

There was no statistically significant effect relationship between peer extraversion and Dutch outcomes in the linear-in-means framework in Table 51, yet in Columns (2) and (3) of Table 58 there is positive and significant relationship between having more peers in the top quintile of the extraversion distribution and Dutch scores. Here the shining light model of having more very extroverted peers to potentially lead the class with their participation might explain these results.

Table 57: Top and Bottom Quintile Model: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score
math_irtscore_t0	0.3426***	0.3404***	0.3421***
	(0.0282)	(0.0277)	(0.0283)
mean_math_irtscore_t0_m	-0.0292		-0.0993
	(0.0784)		(0.0966)
bottom_quintile_math_t0_m		-0.2109	-0.4001
		(0.2819)	(0.3653)
top_quintile_math_t0_m		0.1404	0.1720
		(0.2130)	(0.2395)
getlov_iq	0.0216***	0.0216***	0.0216***
	(0.0025)	(0.0025)	(0.0024)
mean_iq_m	-0.0019		-0.0041
	(0.0067)		(0.0091)
bottom_quintile_getlov_iq_m		0.1993	0.1628
		(0.2480)	(0.2772)
top_quintile_getlov_iq_m		0.2061	0.2234
		(0.1797)	(0.2238)
CON	0.0592***	0.0582***	0.0580***
	(0.0091)	(0.0092)	(0.0092)
mean_CON_m	0.0733***		0.0613*
	(0.0230)		(0.0348)
bottom_quintile_CON_m		-0.3454	-0.0101
		(0.2584)	(0.3440)
top_quintile_CON_m		0.0222	-0.1220
		(0.2530)	(0.2774)
AGREE	0.0057	0.0048	0.0056
	(0.0065)	(0.0064)	(0.0065)
mean_AGREE_m	0.0205		0.0443
	(0.0243)		(0.0423)
bottom_quintile_AGREE_m		0.2690	0.5737
		(0.3080)	(0.4102)
top_quintile_AGREE_m		0.4068*	0.2643
		(0.2263)	(0.2720)
EXTRA	-0.0103	-0.0099	-0.0104
	(0.0071)	(0.0071)	(0.0070)
mean_EXTRA_m	-0.0630**		-0.0560
	(0.0270)		(0.0449)
bottom_quintile_EXTRA_m		0.5161*	0.1946
		(0.2976)	(0.4250)
top_quintile_EXTRA_m		-0.0483	0.1209
		(0.2357)	(0.3099)
Constant	-3.0885***	-3.1219***	-3.2812***
	(0.7389)	(0.3607)	(1.0823)
School FE	Yes	Yes	Yes
Observations	3,255	3,256	3,255
R-squared	0.5552	0.5557	0.5571

Clustered standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income, gender, father's education, and class size not presented in this table

Table 58: Top and Bottom Quintile Model: Dutch Scores

VARIABLES	(1) Dutch score	(2) Dutch score	(3) Dutch score
dutch_irtscore_t0	0.4170***	0.4138***	0.4189***
	(0.0395)	(0.0398)	(0.0396)
mean_dutch_irtscore_t0_d	-0.0252		-0.0354
	(0.0823)		(0.1007)
bottom_quintile_dutch_t0_d		0.1634	0.0786
		(0.1648)	(0.1882)
top_quintile_dutch_t0_d		0.3943***	0.4194**
		(0.1465)	(0.1817)
getlov_iq	0.0169***	0.0167***	0.0169***
	(0.0016)	(0.0016)	(0.0016)
mean_iq_d	-0.0035		-0.0124
	(0.0063)		(0.0103)
bottom_quintile_getlov_iq_d		0.0094	-0.2052
		(0.1795)	(0.2136)
top_quintile_getlov_iq_d		0.2177	0.3945*
		(0.1389)	(0.2156)
CON	0.0469***	0.0468***	0.0462***
	(0.0059)	(0.0058)	(0.0058)
mean_CON_d	0.0577***		0.0587**
	(0.0184)		(0.0236)
bottom_quintile_CON_d		-0.2726*	0.0528
		(0.1560)	(0.2058)
top_quintile_CON_d		-0.2269	-0.3846**
		(0.1689)	(0.1847)
AGREE	0.0005	-0.0001	0.0007
	(0.0040)	(0.0040)	(0.0041)
mean_AGREE_d	-0.0075		-0.0415
	(0.0174)		(0.0312)
bottom_quintile_AGREE_d		0.0780	-0.3067
		(0.2167)	(0.3301)
top_quintile_AGREE_d		0.0964	0.2390
		(0.1561)	(0.1897)
EXTRA	-0.0118**	-0.0117**	-0.0117**
	(0.0047)	(0.0047)	(0.0047)
mean_EXTRA_d	0.0009		-0.0211
	(0.0198)		(0.0320)
bottom_quintile_EXTRA_d		0.0361	-0.0936
		(0.2018)	(0.2836)
top_quintile_EXTRA_d		0.3378*	0.3820*
		(0.1835)	(0.2188)
Constant	-1.8762***	-1.7860***	-0.4601
	(0.6785)	(0.3019)	(1.2264)
School FE	Yes	Yes	Yes
Observations	3,268	3,269	3,268
R-squared	0.7100	0.7121	0.7136

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NB: These regressions also include control variables for peer and individual family income, gender, father's education, and class size not presented in this table

4.8 Conclusion

The analysis in this chapter shows that there is a relationship between peers' non-cognitive characteristics and individual learning outcomes. Estimating a standard linear-in-means model, I find that peer level conscientiousness positively predicts math scores by 0.07 points, while higher peer level extraversion hurts them by -0.06 points. In the case of Dutch scores, I find that peer level conscientiousness positively predicts Dutch scores by 0.06 points. This is the equivalent to a one standard deviation increase in peer level conscientiousness being associated with a 0.14 standard deviation increase in math scores and a 0.129 standard deviation increase in Dutch scores. I generally find no relationship between having higher IQ peers and subject performance, except for pupils above the median sample IQ who benefit more than their peers.

The results of estimating the linear-in-means model at different conditional quantiles prove interesting. Peer level conscientiousness has the strongest relationship with math scores of lower ability pupils, indicating that having diligent and focused peers can pull these pupils up in a significant way. The negative relationship between peer level extraversion and learning outcomes is only statistically significant for high ability pupils. These estimates provide further evidence on the existence of heterogeneous treatment effects.

I find no evidence to support the hypothesis that greater variance in ability within the classroom hurts or benefits learning outcomes, except in the case of peer level conscientiousness and peer level extraversion on Dutch outcomes. In the case of conscientiousness, higher variability has a negative relationship with Dutch performance of the magnitude of -0.01 points, which is the same as a one standard deviation increase in the variance of peer conscientiousness being associated with a -0.032 standard deviation decrease in Dutch scores. For extraversion, the effect is positive, indicating that a higher variance is beneficial to learning outcomes.

I find limited evidence to support a non-linear model of non-cognitive peer effects in the case of math results and more robust evidence in the case of Dutch results. In the math results, having more very agreeable peers positively impacts scores as does having fewer extroverted peers. These results support the shining light model of peer effects. In the Dutch results, I find evidence to support the shining light model of peer effects for past Dutch performance; having more peers from the top quintile of Dutch performance is positively and significantly related to individual outcomes. I also find that having more peers who have low conscientiousness hurts scores and that having more highly conscientious peers hurts scores. This supports the findings from the variance model, which showed that a higher variance of conscientiousness is negatively related to Dutch outcomes. The results for extraversion also support a non-linear model of peer effects. There was no evidence of a peer effect of extraversion in the linear-in-means model; however, the coefficient on having more peers in the top quintile of extraversion is significant and positive.

These results show that peers influence each other's learning outcomes in ways beyond the traditional channels of IQ and past subject performance. Aspects of personality, both at the individual and perhaps more interestingly, at the peer group level, have a meaningful impact on learning outcomes. The finding that higher average peer conscientiousness positively affects math and Dutch outcomes means that schools might want to target interventions to improve pupil conscientiousness. Non-cognitive skills can and should be developed in order to improve both learning and labour market outcomes.

5 Conclusion

The topics addressed in this thesis are inextricably linked. All three chapters broadly ask how education policy can be implemented to positively impact outcomes. From using early education to improve later schooling placement, to extending the school day so mothers have the ability to work, to using information about a pupil's personality when forming classes, education policy is a powerful tool. The connections between education and the labour market mean that all of these types of policies have ramifications beyond the classroom and beyond the individual. Governments can choose to use these types of policies to shape their labour force and thereby shape their economies and societies.

The results of the first chapter showed that early education does not always lead to the desired outcome of better academic results, despite the dominant narrative in the literature and policy-making spheres. This chapter showed that attending an early education programme in Germany for more years did not increase the probability of attending a better secondary school form. In the case of universal programmes, there is limited evidence that they improve learning outcomes on average down the road. This does not mean that early education does not matter, but rather that policymakers need to be clear that the type and quality of the programme will greatly determine what kind of impact it has. Of course academic success is only one of many outcomes and the evidence that such universal programmes positively impact socialisation and other non-cognitive outcomes should not be ignored.

Creating new policies without the necessary information or proper evaluation before a widespread roll-out can lead to unforeseen outcomes. The findings of the second chapter indicate that the extension of the primary school day in Germany has caused non-working mothers to enter the labour market and working mothers to either not change or decrease their hours. For a policy aimed at increasing female labour supply, these results are mixed. Policymakers need to realise that the amount by which they extend the school day will affect a woman's response due to her utility function and the income and substitution effects. An extension of the school day that does not match the full working day will limit mothers' ability to participate full time.

What happens within the classroom ultimately underlies all of the work in education economics. Understanding peer effects, how peers influence each other, allows us to look inside the black box of education. In the third chapter of this thesis, I contribute to the existing literature on peer effects and non-cognitive traits by combining the two and asking how peers' personalities impact learning outcomes using longitudinal data from Flanders, Belgium. The literature on peer effects is mixed, making it difficult to generalise findings. The results in this chapter show that having more conscientious peers improves Dutch and math outcomes, while having more extroverted peers hurts Math outcomes. I find that peer level conscientiousness has the largest impact on Dutch scores of lower performing pupils. In neither case do I find evidence to support the claim that having higher ability peers on average in a classroom improves individual learning outcomes. In the case of Dutch, however, I find evidence pointing to a shining light model of peer effects for past Dutch performance. These results indicate that how pupils learn and interact with each other might ultimately matter more for an individual's academic success than how well their peers perform on average. For school administrators, this means that collecting pupils' personality measures and taking them into account when forming classes could have a positive impact on academic outcomes.

In this thesis, I use detailed data sets and rigorous identification strategies to get at the causal mechanisms underlying three complex issues. There will always be constraints to doing empirical work and despite the limitations posed by the data or context in each chapter, I estimate the most thorough models possible. The goal of this thesis is to contribute to the evidence base on each of these areas of research so that better informed and therefore, more successful policies in the realm of education and the labour market are made.

A Non-cognitive Measures

Table 59 presents the questions from the teacher questionnaire used to construct the non-cognitive measures used in the chapter on non-cognitive peer effects. On this questionnaire, the primary school teacher had to answer how much the statement applied to the pupil using a Likert Scale. These questions were selected from the questionnaire using principal component analysis and then reviewed to assure quality. Each scale was assessed using Cronbach's Alpha to ensure scale reliability (Cronbach, 1951). All measures exceed the rule of thumb value of 0.7 established by Cronbach for a reliable scale.

Table 59: Non-cognitive measures

Original Question Number	Dutch Question	English Translation	Scale	Cronbach's Alpha
Q1	*Kon goed volgen in de klas, heeft voldoende intellectuele mogelijkheden; is verstandig*	Was motivated for school work; wanted to do it really well; worked without reluctance	CON	
Q2	*Was gemotiveerd voor het schoolwerk; wilde het echt goed doen; werkte zonder tegenzin*	Was motivated for school work; wanted to do it really well; worked without reluctance	CON	
Q12	*Kon een samenhangend verhaal vertellen; een onderwerp uitdiepen; bij het onderwerp blijven*	Could tell a coherent story; explore a topic; stay on the subject	CON	
				0.865
Q11	*Stelde zich gemakkelijk open voor de onderwijzer(es); was spontaan; niet defensief*	Was open to the teacher(es); was spontaneous; not defensive	EXTRA	
Q18	*Maakte een energieke en vitale indruk; was vrolijk en zag er gelukkig uit*	Made an energetic and vital impression; was smiling and looked happy	EXTRA	
Q21	*Zocht contact met de medeleerlingen; was open en aanspreekbaar*	Made contact with fellow students; was open and approachable	EXTRA	
				0.793
Q5	*Stoorde de les niet opzettelijk; was niet gericht op het boycotten van het lesverloop*	Did not disturb the lesson intensionally; did not aim to boycott learning	AGREE	
Q9	*Hield zich goed (dit is uit zichzelf) aan de klasregels; wachtte zijn/haar beurt af; voortdurend tot de orde roepen was niet nodig*	Held herself to the class rules; waited for her turn; it was not necessary to constantly call her to order	AGREE	
Q15	*Was afkerig van vijandelijkheden; wilde vriendelijk en aardig zijn voor anderen; beleefde geen genoegen aan het plagen en pesten van anderen*	Was averse to hostilities; was friendly and kind to others; experienced no pleasure in teasing and bullying of others	AGREE	
				0.843

Source: LOSO Data Set ("Beoordeling van der leerling door de leerkracht Basisonderwijs")

References

[1] Aizer, Anna. Peer effects and human capital accumulation: The externalities of ADD. No. w14354. National Bureau of Economic Research, 2008.

[2] Almlund, Mathilde, et al. "Personality Psychology and Economics." *Handbook of the Economics of Education* 4.1 (2011).

[3] Ammermueller, Andreas, and Jörn-Steffen Pischke. "Peer effects in European primary schools: Evidence from the progress in international reading literacy study." *Journal of Labor Economics* 27.3 (2009): 315-348.

[4] Baker, Michael, Jonathan Gruber, and Kevin Milligan. "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *Journal of Political Economy*, University of Chicago Press, 116.4 (2008): 709-745.

[5] Bargar, Robert R., and Randy L. Hoover. "Psychological type and the matching of cognitive styles." *Theory into Practice* 23.1 (1984): 56-63.

[6] BBC. "Languages." BBC News. BBC, 2014. Web. 16 Apr. 2015.

[7] Becker, Gary Stanley. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Chicago: University of Chicago, 1983.

[8] Berlinski, Samuel, Sebastian Galiani, and Paul Gertler. "The effect of pre-primary education on primary school performance." Journal of Public Economics 93.1 (2009): 219-234.

[9] Bertrand, Marianne, and Jessica Pan. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal: Applied Economics*, American Economic Association, vol. 5.1 (2013): 32-64.

[10] Blanden, Jo, et al. "Universal pre-school education: the case of public funding with private provision," CEP Discussion Paper No 1352, (2015).

[11] Blau, David, and Janet Currie. "Pre-school, day care, and after-school care: who's minding the kids?." Handbook of the Economics of Education 2 (2006): 1163-1278.

[12] Borg, Mary O., and Stephen L. Shapiro. "Personality type and student performance in principles of economics." *The Journal of Economic Education* 27.1 (1996): 3-25.

[13] Borghans, Lex, et al. "The economics and psychology of personality traits." *Journal of Human Resources* 43.4 (2008): 972-1059.

[14] Bound, John, David A. Jaeger, and Regina M. Baker. "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 (1995): 443-450.

[15] Bowles, Samuel, Herbert Gintis, and Melissa Osborne. "The determinants of earnings: A behavioral approach." *Journal of Economic Literature* (2001): 1137-1176.

[16] Brewer, Mike, and Claire Crawford. Starting school and leaving welfare: the impact of public education on lone parents' welfare receipt. No. 10, 19. IFS Working Papers, 2010.

[17] Burke, Mary A., and Tim R. Sass. "Classroom peer effects and student achievement." *Journal of Labor Economics* 31.1 (2013): 51-82.

[18] Cameron, Adrian Colin, and P. K. Trivedi. *Microeconometrics: Methods and Applications.* Cambridge: Cambridge UP, 2005.

[19] Campbell, F.A. and C. T. Ramey. "Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families." *Child Development*, 65.2 (1994): 684–698.

[20] Campbell, Frances, et al. "Early childhood investments substantially boost adult health." *Science* 343.6178 (2014): 1478-1485.

[21] Campolieti, Michele, and Chris Riddell. "Disability policy and the labor market: Evidence from a natural experiment in Canada, 1998–2006." Journal of Public Economics 96.3 (2012): 306-316.

[22] Card, David. "The causal effect of education on earnings." *Handbook of Labor Economics.* Vol. 3 (1999): 1801-1863.

[23] Card, David and Krueger, Alan. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania", *American Economic Review*, 84.4 (1994): 772-793.

[24] Cascio, Elizabeth U., and Diane Whitmore Schanzenbach. The impacts of expanding access to high-quality preschool education. No. w19735. National Bureau of Economic Research, 2013.

[25] Costa, Paul T., and Robert R. McCrae. "Four ways five factors are basic." *Personality and individual differences* 13.6 (1992): 653-665.

[26] Contreras, Dante, Paulina Sepúlveda, and Soledad Cabrera. "The effects of lengthening the school day on female labor supply: Evidence from a quasi-experiment in Chile." Centro de Microdatos, Departamento de Economía. Universidad de Chile (2010).

[27] Cronbach, Lee J. "Coefficient alpha and the internal structure of tests." *psychometrika* 16.3 (1951): 297-334.

[28] Cunha, Flavio and Heckman, James J. "The Technology of Skill Formation," *American Economic Review.* 97.2 (2007): 31-47.

[29] Currie, Janet. "Early Childhood Education Programs." *The Journal of Economic Perspectives.* (2001): 213-238

[30] Currie, Janet, and Duncan Thomas. "Does Head Start Make a Difference?". *American Economic Review.* 85.3 (1995): 341-364.

[31] DeMars, Christine. *Item response theory.* Oxford University Press, 2010.

[32] Dominiczak, Peter. "David Cameron: Working Parents to save £5,000 a Year under Tory Childcare Pledge." The Telegraph. Telegraph Media Group, 14 Apr. 2015. Web. 19 Aug. 2015.

[33] Dustmann, Christian, Anna Raute, and Uta Schönberg. "Does universal child care matter? Evidence from a large expansion in pre-school education." unpublished manuscript, (2012).

[34] Federal Reserve Bank of St. Louis. "Belgium / U.S. Foreign Exchange Rate." 8 Mar. 2006. Web. 03 July 2015.

[35] Felfe, Christina, and Rafael Lalive. "How Does Early Child Care affect Child Development?" mimeo. (2011). Working Paper.

[36] Fertig, Michael, and Jochen Kluve. "The Effect of Age at School Entry on Educational Achievement in Germany." IZA Discussion Paper 1507 (2005).

[37] Fraine, Bieke, et al. "The effect of schools and classes on language achievement." *British Educational Research Journal* 29.6 (2003): 841-859.

[38] Fuerst, J. S. and Dorothy Fuerst. "Chicago Experience with an Early Childhood Program: The Special Case of the Child Parent Center Program." *Urban Education.* 28 (1993): 69-96.

[39] Garces, Eliana, Duncan Thomas, and Janet Currie. "Longer-Term Effects Of Head Start," *American Economic Review.* (2002): 999-1012.

[40] Geay, Charlotte, Sandra McNally, and Shqiponja Telhaj. "Non-native Speakers of English in the Classroom: What Are the Effects on Pupil Performance?" *The Economic Journal* 123.570 (2013): F281-F307.

[41] Gelbach, Jonah B. "Public schooling for young children and maternal labor supply." *American Economic Review* (2002): 307-322.

[42] Georgii, Harald. "Rechtsanspruch auf einen Kindergartenplatz." *Neue Juristische Wochenzeitschrift* (1996): 686-688.

[43] Germany. Statistisches Bundesamt. Statistik Der Kindertagesbetreuung. N.p., 2014. Web. 13 Nov. 2014.

[44] Germany. Statistisches Bundesamt. Allgemeinbildende und berufliche Schulen: Lehrkräfte insgesamt sowie Anteil der weiblichen Lehrkräfte nach Schularten und Beschäftigungsumfang. N.p., 2014. Web. 3 Dec. 2014.

[45] Gibbons, Stephen, and Shqiponja Telhaj. "Peer effects: evidence from secondary school transition in England." *Oxford Bulletin of Economics and Statistics* (2015).

[46] Goldberg, Lewis R. "A Historical Survey of Personality Scales and Inventories." In *Advances in Psychological Assessment*, ed. Paul McReynolds. (1971): 293-336.

[47] Goldberg, Lewis R. "An Alternative Description of Personality: The Big-Five Factor Structure." *Journal of Personality and Social Psychology* 59.6 (1990): 1216.

[48] Hamermesh, Daniel S., and Stephen J. Trejo. "The demand for hours of labor: Direct evidence from California." Review of Economics and Statistics 82.1 (2000): 38-47.

[49] Heckman, James. "Effects of Child-Care Programs on Women's Work Effort." Journal of Political Economy 82.2 (1974): S136-S163.

[50] Heckman, James J., and Yona Rubinstein. "The importance of noncognitive skills: Lessons from the GED testing program." *American Economic Review* (2001): 145-149.

[51] Heckman, James J., Lance J. Lochner, and Petra E. Todd. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," *Handbook of the Economics of Education*. Elsevier, 2006.

[52] Heckman, James J., Jora Stixrud, and Sergio Urzua. "The Effects Of Cognitive and Noncognitive Abilities On Labor Market Outcomes and Social Behavior," *Journal of Labor Economics* 24.3 (2006): 411-482.

[53] Heckman, James J., and Tim Kautz. "Hard evidence on soft skills." *Labour economics* 19.4 (2012): 451-464.

[54] Heckman, James J., Roberto Pinto, and Peter Savelyev. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103.6 (2013): 1-35.

[55] Hinshaw, Stephen P. "Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms." *Psychological bulletin* 111.1 (1992): 127.

[56] Holtappels, Heinz Günter. Ganztagsschule in Deutschland: Ergebnisse Der Ausgangserhebung Der "Studie Zur Entwicklung Von Ganztagsschulen" (StEG). Ed. Eckhard Klieme, Thomas Rauschenbach, and Ludwig Stecher. Weinheim: Juventa Verl., 2008. Print.

[57] Hunt, Jennifer. "Why do People Still Live In East Germany?," DIW Discussion Paper, No. 201, (2000) DIW Berlin.

[58] International Labor Organisation Statistics (ILOStat). "Employment-to-population ratio by sex and age (%)." <http://www.ilo.org/ilostat>

[59] John, Oliver P., and Sanjay Srivastava. "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." *Handbook of Personality: Theory and Research* 2.1999 (1999): 102-138.

[60] Kellogg, Ryan, and Hendrik Wolff. "Daylight time and energy: Evidence from an Australian experiment." Journal of Environmental Economics and Management 56.3 (2008): 207-220.

[61] Koenker, Roger, and Kevin Hallock. "Quantile regression: An introduction." *Journal of Economic Perspectives* 15.4 (2001): 43-56.

[62] Lancksweerdt, P. "Getlov (Gemeenschappelijke Testbatterij Lager Onderwijs)." Onderzoeksverslag over de gegevens van de leerlingen van het 6de leerjaar lager onderwijs uit de vrije PMS-centra in Oost-Vlaanderen schooljaar 1989-1990 (1991).

[63] Lefebvre, Pierre, and Philip Merrigan. "Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada." *Journal of Labor Economics* 26.3 (2008): 519-548.

[64] Machin, Stephen, Olivier Marie, and Sunčica Vujić. "The Crime Reducing Effect of Education." *The Economic Journal* 121.552 (2011): 463-484.

[65] Manski, Charles F. "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies* 60.3 (1993): 531-542.

[66] Marcus, Jan, Janina Nemitz, and C. Katharina Spieß. "Ausbau der Ganztagsschule: Kinder aus einkommensschwachen Haushalten im Westen nutzen Angebote verstärkt" *DIW Wochenbericht*, 27(2013): 11-23.

[67] Meghir, Costas, and Marten Palme. "Educational Reform, Ability, and Family Background." *American Economic Review.* 95.1 (2005): 414-24.

[68] Mundlak, Yair. "On the Pooling of Time Series and Cross Section Data." *Econometric*a 46.1 (1978): 69-85.

[69] Organisation for Economic Cooperation and Development. “Messages From PISA 2000”. (2004): 41-75.

[70] OECD, “How do girls compare to boys in mathematics skills?”, In *PISA 2009 at a Glance (2011)*, OECD Publishing.

[71] OECD, *School Evaluation in the Flemish Community of Belgium 2011, OECD Reviews of Evaluation and Assessment in Education (2011)*, OECD Publishing. http://dx.doi.org/10.1787/9789264116726-en

[72] OECD, "BELGIUM–Country Note–Results from PISA 2012." (2013):6. OECD Publishing. http://www.oecd.org/belgium/PISA-2012-results-belgium.pdf

[73] "Ofsted - GOV.UK." Ofsted - GOV.UK. N.p., n.d. Web. 08 Feb. 2016.

[74] "Ofsted School Inspection Handbook." Ofsted - GOV.UK. 2015. Web. 08 Feb. 2016.

[75] Opdenakker, Marie-Christine, et al. "The effect of schools and classes on mathematics achievement." *School effectiveness and school improvement* 13.4 (2002): 399-427.

[76] Poropat, Arthur E. "A meta-analysis of the five-factor model of personality and academic performance." *Psychological bulletin* 135.2 (2009): 322.

[77] Prasad, Eswar S. "The Unbearable Stability of the German Wage Structure: Evidence and Interpretation," IMF Staff Papers, Palgrave Macmillan, Vol. 51.2 (2004): 354-385.

[78] Puhani, Patrick A. and Andrea M. Weber. “Does the early bird catch the worm? Instrumental variable estimates of early education effects of age of school entry in Germany.” *Empirical Economics*. (2007): 359-386.

[79] Rainer, Helmut, et al. "Kinderbetreuung." *ifo Forschungsberichte* 59 (2013).

[80] Riedel, Birgit. "Das institutionelle Angebot für Kinder ab 6 Jahren (Grundschulalter)." In *Deutsches Jugendinstitut eV/Dortmunder Arbeitsstelle Kinder-und Jugendhilfestatistik: Zahlenspiegel* (2005): 143-155.

[81] Robelen, Erik W. "The Great Divide." *Education Week* (2005): 31-35.

[82] Rothstein, Jesse. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, vol. 125.1(2010): 175-214.

[83] Sacerdote, Bruce. "Peer effects in education: How might they work, how big are they and how much do we know thus far?." *Handbook of the Economics of Education*, Vol.3 (2011): 249-277.

[84] Schweinhart, Lawrence J., Helen Barnes and David Weikart. *Significant Benefits: The High/Scope Perry Preschool Study Through Age 27. Monograph of the High/Scope Educational Research Foundation.* Ypsilanti, Michigan: High-Scope Educational Research Foundation (1993).

[85] Sirin, Selcuk R. "Socioeconomic status and academic achievement: A meta-analytic review of research." *Review of educational research* 75.3 (2005): 417-453.

[86] Slesnick, Daniel T. "Empirical approaches to the measurement of welfare." *Journal of Economic Literature* 36.4 (1998): 2108-2165.

[87] Socio-economic Panel (SOEP) (2010), Data for years 1984-2009, Version 26, SOEP.

[88] Socio-economic Panel (SOEP) (2013), Data for years 1984-2012, Version 29, SOEP.

[89] Spiess, C. Katharina, Felix Büchel, and Gert G. Wagner. "Children's School Placement in Germany: Does Kindergarten Attendance Matter?" IZA Discussion Paper (2003).

[90] Spiess, C. Katharina. "Betreuungsgeld widerspricht den zielen nachhaltiger familienpolitik." *DIW Wochenbericht*, 79.24 (2012).

[91] Staiger, Douglas and James H. Stock. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65.3 (1997): 557-586.

[92] Statistisches Bundesamt. DESTATIS. "Schulen Auf Einen Blick." Statistisches Bundesamt, Wiesbaden, Mar. 2012. Web.

[93] Quinn, Simon. "Bank Credit and Legal Status in Moroccan Manufacturing." DPhil Thesis, University of Oxford. (2010).

[94] Van Damme, Jan, et al. "A new study on educational effectiveness in secondary schools in Flanders: An introduction." *School Effectiveness and School Improvement* 13.4 (2002): 383-397.

[95] Van Landeghem, Georges, et al. "The effect of schools and classes on noncognitive outcomes." *School Effectiveness and School Improvement* 13.4 (2002): 429-451.

[96] Vigdor, Jacob, and Thomas Nechyba. *Peer effects in North Carolina public schools. Schools and the Equal Opportunity Problem*, MIT Press, 2007.

[97] Vytlacil, Edward. "Independence, monotonicity, and latent index models: An equivalence result." *Econometrica* 70.1 (2002): 331–341.

[98] "Wege Zum Abschluss - Viele Möglichkeiten Bei Weiterführenden Schulen." Wege Zum Abschluss - Viele Möglichkeiten Bei Weiterführenden Schulen. 3sat, 31 May 2011. Web. 15 Aug. 2015.

[99] Wenzel, Stefanie. "Konvergenz oder Divergenz? Einstellungen zur Erwerbstätigkeit von Müttern in Ost- und Westdeutschland." *GENDER–Zeitschrift für Geschlecht, Kultur und Gesellschaft* 2.3 (2010).

[100] Wittrock, Philipp. "Kita Oder Klage." *Der Spiegel.* N.p., 30 May 2012. Web. 13 Aug. 2012.

[101] Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT, 2002.