ifo
BEITRÄGE
zur Wirtschaftsforschung

76
2017

Economic Perspectives on the Implications of Public Child Care and Schooling for Educational Outcomes in Childhood and Adult Life

Larissa Zierow

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

ifo
BEITRÄGE
zur Wirtschaftsforschung

76
2017

Economic Perspectives on the Implications of Public Child Care and Schooling for Educational Outcomes in Childhood and Adult Life

Larissa Zierow

Herausgeber der Reihe: Clemens Fuest
Schriftleitung: Chang Woon Nam

ifo INSTITUT
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische
Daten sind im Internet über
http://dnb.d-nb.de
abrufbar.

ISBN: 978-3-95942-034-1

Druck: ifo Institut, München

ifo Institut im Internet:
http://www.cesifo-group.de

Preface

This thesis was written by Larissa Zierow while having been a research assistant at the Seminar for Population Economics at the University of Munich. It was completed in December 2016 and accepted as a doctoral thesis by the Department of Economics at the University of Munich in May 2017.

Child care and education are important inputs in the human capital production function. The research of Nobel prize winner James Heckman shows that skills are built from the early childhood on and increase the returns of later educational inputs, in short: *skill begets skill*. Therefore, it is crucial for a society to know how the public education system can support the skill development of children and thus its labor force's skills of tomorrow. This thesis aims to provide knowledge on the effects of some important aspects of education policies. It comprises an introduction and four core chapters which cover microeconometric evaluations of child care and schooling reforms in Germany using unique administrative data and rich survey data. These empirical evaluations contribute to three understudied fields in the literature on individual returns to education: first, the intensive margin of child care, second, the quality of child care, and finally, denominational sorting and classroom composition.

Chapter 2 evaluates the impact of after-school center-based care on the development of primary school-aged children. The analysis is based on detailed data of the German Child Panel and employs a value-added method. While we do not find significant effects on average, our analysis provides evidence for beneficial returns to after-school center-based care attendance for more disadvantaged children.

Chapter 3 investigates the consequences of expanding the supply of child care centers operating on a full-day basis on children's skills just before entering primary school. Identification relies on a substantial expansion of the number of full-day slots triggered by several reforms of the German child care system. Employing a municipality fixed-effect approach and using unique administrative data covering almost 100,000

children in the German state of Schleswig-Holstein, we find negative effects on children's socio-emotional maturity. Nevertheless, immigrant children benefit from the expansion of full-day care in terms of school readiness.

Chapter 4 investigates the effects of quality reforms on early child care supply and on child development outcomes. The analysis is based on administrative data covering all child care centers and family day carers in Germany and school entrance examination data of North Rhine-Westphalia, a large German state. The results of a region fixed-effect estimation framework show that stricter regulations of staff quality have significant negative average effects on the supply of care. Moreover, poorer counties are more likely to expand supply by increasing their share of family day care, which is a less expensive care mode than center-based care, but is characterized by lower staff qualification standards. In contrast to the existing literature, the results show that a higher share of family day care is not detrimental for children's development.

Chapter 5 analyzes the effects of the abolition of denominational schools in the Saarland, a German state. Such schools are an important education provider in many countries around the world. Due to their focus, these schools often operate with multigrade classes, in which more than one age cohort is taught in one classroom. Multigrade classes are a cost-effective way to provide education and play a crucial role in education policy in the context of demographic change. We employ a triple difference-in-difference approach to estimate the causal effect of attending denominational schools with multigrade classes on schooling and short-run labor market outcomes. The findings document positive effects of switching from a multigrade to a single-grade school system on students' final grade attainment and their labor market participation.

Keywords: child care, child development, child care quality, regulation, education, multigrade schools.

JEL-Codes: J13, I21, I26, I28, H75, L51.

Acknowledgments

I am deeply indebted to all those who supported me during the past five years when this dissertation was written.

First and foremost, I am particularly grateful to my supervisor Uwe Sunde for his continuous support, encouragement and invaluable advice. I would like to thank him for his confidence in my skills and for motivating me to start an academic career. Second, I would like to thank Ludger Wössmann for co-supervising my thesis, as well as for providing me with valuable feedback along my way as PhD student. Third, special gratitude goes to Christina Felfe, with whom I have had the privilege of working together on two projects. She has supported me also beyond our joint work with helpful advice and constructive suggestions. Fourth, I would like to thank my colleagues, fellow PhD students, administrative staff and friends at the University of Munich, at the Ifo Institute for Economic Research and at the University of St. Gallen for valuable comments, fruitful discussions, and also for their personal support. I owe special thanks to Rudi Stracke, Lukas Buchheim, Ilka Gerhardts, Ruth Schüler, Thomas Vischer, Beatrice Mäder, Nadjeschda Arnold and Susan Fay. Fifth, I would not have been able to write this thesis without the support of great people working in German ministries and statistical offices who believed in the added value of my projects and provided access to various administrative data sources: I would like to thank Heidrun Thaiss, Brigitte Borrmann, Klaus Simon, Heiko Bergmann, Nina Storfinger and Thomas May for their help and hugely useful comments on my projects.

Finally, I would like to thank my partner Michal and my whole family for always having shown interest in my work and for their practical and emotional support throughout the years. Special thanks go to my little son Jan Vincent for filling me with joy and happiness every day.
I dedicate this book to my parents, Inge and Kalle, who taught me to value education.

Economic Perspectives
on the
Implications of Public Child Care and Schooling
for
Educational Outcomes in Childhood and Adult Life

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae
(Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2016

vorgelegt von

Larissa Zierow

Referent: Prof. Dr. Uwe Sunde
Korreferent: Prof. Dr. Ludger Wößmann

Promotionsabschlussberatung: 10. Mai 2017

Contents

List of Figures

List of Tables

Teach the children so it will not be necessary to teach the adults.
Abraham Lincoln

Chapter 1

General Introduction

1.1 An Economist's Perspective on Public Child Care and Education

There are many non-economic reasons for investing in public child care and education. Yet, there are also important economic reasons. In economic theory, skills and knowledge of an individual are subsumed as her human capital, a definition that was coined by the work of Becker (1964). Education is an investment in the accumulation of human capital and plays an important role in all standard theories that aim to explain individual labor market outcomes as well as macroeconomic outcomes. Empirical studies show that education enhances individuals' productivity and reduces their risk of unemployment.[1] At the macroeconomic level, education has been shown to exert positive effects on the economic standing of a country and its economic growth, also in the very long run.[2] Importantly, human capital does not only subsume cognitive skills such as an individual's IQ, her highest degree or PISA[3] test score, but also her non-cognitive skills. Self-control, motivation, self-esteem, concentrativeness and endurance are highly correlated with school attendance and educational success. Furthermore, non-cognitive skills are negatively associated with

[1] See Heckman *et al.* (2006a) and Psacharopoulos (1994) for an overview.
[2] Overviews of studies on the macroeconomic effects of education are provided by Woessmann (2016), Hanushek and Woessmann (2008), Sunde and Vischer (2015), and Galor (2005).
[3] OECD's Programme for International Student Assessment

smoking and crime, and positively associated with income, wealth and physical as well as psychological health.[4]

Economists have always been interested in understanding the sources of inequality between and within countries – starting with the so-called founder of economic thought Adam Smith (Smith, 1904) to today's bestseller on economic inequality by Piketty (2014). Not only capital, but also human capital play a role in explaining inequality. The recent literature, headed by James Heckman, shows that differences in skills and competencies between individuals open up early in life. The conditions before birth and in the first years of life have important consequences for the individual economic success and for the economic standing of a country as a whole (Cunha *et al.*, 2006). Following this line of argument, child care and early education can be seen as an instrument that helps reducing inequality within society.[5] At the same time, education and care for the youngest part of the population are not only tools to invest in human capital from the very beginning, but are also part of family politics. Public child care aims at facilitating the combination of professional and family life, especially by enabling more mothers to engage in the labor market. Economists are thus not only interested in the effects of early education and care on children's skill development but also in its effect on maternal labor supply.

Education systems around the world are regularly subject to reforms. In the last years, especially early child care and education gained importance and many countries undertook reforms in this area (OECD, 2016). In order to find out whether the reforms were effective and led to the desired outcomes, or whether further changes are required, it is important to evaluate the reforms. This thesis evaluates educational reforms in the German education system. In Chapters 2-4, the effects of after-school care and the effects of the reform-induced expansion of universal child care on individual cognitive and non-cognitive skills are studied. In Chapter 5, the effects of the

[4] Moffitt *et al.* (2011), Heckman *et al.* (2006b), Chamorro-Premuzic and Furnham (2006, 2005), McLeod and Kaiser (2004), and Bowles *et al.* (2001) provide, amongst others, evidence for the effects of non-cognitive skills on educational success and other outcomes. Moreover, recent studies also focus on the interplay between non-cognitive and cognitive skills. Dohmen *et al.* (2010), for example, show that cognitive skills are negatively correlated with risky behavior.

[5] See evidence provided by, amongst others, Magnuson and Duncan (2016), Heckman (2011), Almond and Currie (2011) and Janet Currie (1999).

legal abolition of denominational schools on individual schooling and labor market participation are investigated.

1.2 The Case of Germany: Reforms of the Educational System

Germany represents an interesting case to study the effects of child care policies. Only recently, major reforms were undertaken to overcome traditional gender stereotypes and facilitate the combination of professional and family life for both fathers and mothers. Nevertheless, in comparison to other countries, there is still a lot to be done: according to OECD (2011), in no other country it is so difficult for families to combine work and family life. Compared to the US and to other European countries (e.g. Sweden and France), mothers stay at home after birth and are absent from the labor market for a much longer time. This characteristic has reasons to be found in the past: in the 1970s, the maximum duration of parental leave had been increased from 6 months to 3 years. This reform aimed at providing more equal opportunities for women in the labor market, but in fact, given the simultaneous low availability of public child care, it was an incentive for many mothers to stay at home. Furthermore, the tax system in Germany supports the traditional bread winner model by allowing for *Ehegattensplitting* (tax system in which husband and wife each pay income tax on half the total of their combined incomes). In consequence, for a very long time, it was common that young children stayed at home with their mothers, at least until they reached kindergarten age (i.e. turned 3 years old). A major reform from 1996 implemented a legal entitlement for every child turning 3 years to a slot in public child care. As a result of this policy, since the early 2000s, more than 90% of children entering school have attended kindergarten for at least 2 years. In 2005 and 2008, politicians announced a legal entitlement to a slot in early child care from 2013 on. In the following years, a large expansion of early child care has been observed, but the availability is still very low. In 2012, early child care coverage was at 28%. Yet, 40% of parents stated that they would like to enroll their child in early care (BMFSFJ,

2014). Reforms in 2005 and 2008 also contained new rules on expanding full-day care for children. Until then, half-day kindergarten and half-day primary school was the standard, which was a major reason for the low full-time employment share amongst mothers. The reforms led to an expansion in full-day slots in kindergarten and after-school care for children in primary school.

Inequality is another interesting aspect for using Germany as region of study. According to the PISA results in 2000, individual success at school in Germany is more dependent on parental background than in any other participating country and there is a large regional variation regarding equal opportunities provided by the education system (Baumert *et al.*, 2001; Berkemeyer *et al.*, 2014). Another educational reform evaluated in this thesis took place in the German state *Saarland* in 1969, having impacted inequality between rural and urban students. The reform led to the abolition of denominational schools (Catholic and Protestant students had been taught in different buildings by different teachers until then). In rural regions with low population density, denominational schools – due to their focus – often had to operate with multigrade classes, in which more than one age cohort is taught in one classroom. On the other hand, children living in urban areas were taught on a single-grade basis as the population density in their regions allowed separating classes by age even when splitting up students by denomination. The abolition of the reform in 1969 thus led to a harmonization of learning conditions of rural and urban student populations.

1.3 Outline

This thesis consists of four contributions to the fields of economics of education and early child care economics. Each contribution corresponds to one chapter; all chapters are self-contained and can be read independently.

Chapter 2 (in collaboration with Christina Felfe) evaluates the impact of after-school center-based care on the development of primary school-aged children. The estima-

tion of this impact is challenging due to the non-random selection of children into after-school center-based care. We tackle this challenge by using detailed data of the German Child Panel and employing a value-added method. While we do not find significant effects on average, our analysis provides evidence for beneficial returns to after-school center-based care attendance for more disadvantaged children. To be more precise, children of less-educated mothers and low-income families benefit from attending after-school care centers in terms of their socio-behavioral development.

Chapter 3 (in collaboration with Christina Felfe) investigates the consequences of expanding the supply of child care centers operating on a full-day basis on children's skills just before entering primary school. Identification relies on a substantial expansion of the number of full-day slots triggered by several reforms of the German child care system. Using unique administrative data covering almost 100,000 children in the German state of Schleswig-Holstein, we find negative effects on children's socio-emotional maturity. These effects are driven by children from disadvantaged backgrounds, in particular, children of single parents, low-educated families and immigrant ancestry. Nevertheless, immigrant children benefit from the expansion of full-day care in terms of school readiness.

Chapter 4 uses reforms of child care quantity and quality in Germany to investigate questions that have not been studied in combination so far: the effects of the legal and economic environment on public early child care quantity and quality and the effects of the resulting early child care quantity and quality on child development outcomes. The analysis is based on administrative data covering all child care centers and family day carers in Germany and school entrance examination data of the German state of North Rhine-Westphalia. I find that regulations of staff quality have significant negative average effects on the quantity of provided care. Moreover, the results show that poorer counties are more likely to expand their early child care supply by increasing their share of family day care, which is a less expensive and more flexible model than center-based care, but is characterized by lower staff qualification standards. In contrast to the existing literature, the results show that a higher share of family day care is not detrimental for children's development. Instead, a larger share

of family day care – when accompanied by care quality regulations – has positive effects on the development of children's socio-emotional stability and motor skills.

Chapter 5 (in collaboration with Ilka Gerhardts and Uwe Sunde) analyzes the effects of the abolition of denominational schools in the Saarland, a German state. Such schools are an important education provider in many countries around the world. Due to their focus, these schools often operate with multigrade classes, in which more than one age cohort is taught in one classroom. Multigrade classes are a cost-effective way to provide education and play a crucial role in education policy in the context of the demographic change. We present estimates of the causal effect of attending denominational schools with multigrade classes on schooling and short-run labor market outcomes. The analysis exploits the abolition of denominational schools in the Saarland in 1969 for the identification of the effect. The findings document positive effects of switching from a multigrade to a single-grade school system on students' final grade attainment and their labor market participation. Notably, the positive impact is most pronounced in the outcomes of girls.

1.4 Related Literature and Contributions of this Thesis

The first economic literature on effects of child care was mainly referring to effects of targeted care, such as the Perry Preschool Program in the US (Heckman *et al.*, 2010). These programs are targeted at children from rather disadvantaged family backgrounds. Furthermore, their quality of care is very high and directly aligned with the needs of the participating children. Consequently, the effects of such programs are likely to present an upper bound of potential effects of universal child care (Blau and Currie, 2006).

The more recent literature focuses on effects of universal child care programs to which every child has access and thus, there is a large variation regarding parental background. Baker *et al.* (2008) find that the introduction of universal child care in

Quebec had negative effects, while other studies (focusing on other countries) find that the expansion of universal child care had positive effects (e.g. Havnes and Mogstad (2011) in Norway and Cornelissen *et al.* (2016) in Germany). What the economic literature on effects of universal child care did not take into account so far is the intensive margin of care. In other words, much attention was paid to child care attendance per se but not to the amount of hours spent in child care. Yet, the effect of longer hours in care is of interest since full-day child care becomes more prevalent in many countries. In Chapter 2, we look at the effects of care provided in the afternoon on primary school children. In Chapter 3, we study the effects of providing child care for kindergarten-aged children on a full-day basis. In both cases, attending full-day care leads to a reduction of time spent with the primary care giver, in most cases the mother. Effects accordingly vary across family background characteristics since the counterfactual care varies with respect to the education of the mother, the ability of parenting etc. We find in both studies that more hours in care make a difference. While immigrant children benefit in terms of cognitive development both from after-school care attendance and from full-day kindergarten, overall socio-emotional stability of kindergarten-aged children is negatively affected when attending full-day care. In both studies, we do not find any significant effect for children from more advantaged, high-educated backgrounds.

Another important aspect when studying effects of universal child care is the quality of care. Targeted programs – such as the Perry Preschool Program – are usually of a very high quality which makes these programs quite expensive (and often not scalable). Does universal child care have positive effects on children even if the quality of care is not as high? Coming along with this question is the more basic question: from the perspective of public policies, what is the most effective way to achieve quality in child care on a large scale? From an economist's perspective, the quality of goods is regulated by markets, but only in the context of experience goods. In education and child care, it is hard to judge the quality upfront due to asymmetric information (Akerlof, 1970). Therefore, there is a need for regulations on quality in education and care. As Blau (2007) and Hotz and Xiao (2011) show, however, regulations can have unintended effects. In the US, stricter quality rules in formal

child care led to a decrease of formal child care in low-income regions and to an increase of informal child care (of lower quality) in these regions. In Chapter 4 of this thesis, I study the effects of child care regulations in Germany. First, I analyze how the economic and legal environment affects the child care market and the quality of the supplied child care. Second, I study the effects of an increased family-day care share – a care model which is characterized by a lower pedagogical quality of staff – on child outcomes.

Peer effects are further central aspects of education research. They are regarded as important inputs to the education production function ever since Coleman (1968) made them popular (see for example Iversen and Bonesrønning (2015) and Jones (2013)). One setting that produces multiple forms of peer effects are multigrade classes – a teaching form that we study in the Chapter 5 of this thesis. Peer effects may be positive if more skilled classmates serve as natural role models (Duflo *et al.*, 2011; Hanushek *et al.*, 2003). Peer effects may also be negative, for example when age gaps arise due to grade repeating and redshirting (Lavy *et al.*, 2012; Jones, 2013). Our study contributes to the literature by providing evidence of class composition effects that are placed in a Western European society, whereas existing studies on multigrade classes have been mainly focused on developing countries. The use of the natural experiment of the sudden abolition of denominational schools allows for a credible identification of the causal impact of denominational schools with multigrade classes.

1.5 Data

Meaningful evaluation of reforms in the educational system hinges upon available data. In Germany, data availability is not as high as, for example, in Scandinavian countries in which evaluations on long-run effects are more easily feasible since it is possible to link individual schooling variables with later labor market participation and income via a personal ID (see for example Havnes and Mogstad (2011) on long run effects of child care). In Germany, data protection rules are quite strict and there is no personal ID available to link such outcomes. Nevertheless, in the last decade more

data – both survey data and administrative data – became publicly available that make it possible to evaluate reforms. There is a trade-off when deciding for German administrative data or survey data when doing empirical evaluations: administrative data offer information on the full population and thus do not suffer from small sample and response bias as survey data. Yet, data on treatment and outcome of interest are not always possible to link at the individual level – due to data protection rules – so that effects can only be interpreted as intention-to-treat effects. The chapters of this thesis present results based on survey data as well as on administrative data and respectively explain the advantages and limits of their type.

The data set used in Chapter 2 is the German Child Panel (GCP). The GCP is a longitudinal survey conducted by the German Youth Institute from 2002 to 2005. Besides providing access to a wide range of information on important background characteristics, the GCP contains a broad spectrum of indicators on child development (such as children's school grades and measures of children's socio-behavioral development) and child care provision (such as center-based after-school care attendance). A second dataset also based on surveys in families is called *Familie in Deutschland* (FiD). This dataset is provided by the *Deutsches Institut für Wirtschaftsforschung* (DIW). It includes birth cohorts from 2004 to 2007 and provides information on children's socio-emotional maturity as well as children's full-day care attendance. This dataset is used in Chapter 3 of this thesis to provide additional evidence on mechanisms driving the effects of full-day child care.

Data used for the main analysis in Chapter 3 and Chapter 4 of this thesis are administrative records of all child care institutions in Germany for the years 1998-2011: the so-called *Kinder- und Jugendhilfestatistik* or *Statistics on Child and Youth Services*. They do not only include information on the number of children and staff in each institution but also detailed information on structural quality indicators such as group size, education of staff and staff-to-child ratio.

Further administrative data used for evaluations undertaken in this thesis are administrative records from *Schleswig-Holstein*'s and *North Rhine-Westphalia*'s school entrance examinations (SEE) drawing on data for the school entrance cohorts 2004 to 2012

(Chapter 3 of this thesis) and school entrance cohorts 2010 to 2013 (Chapter 4 of this thesis). SEE data contain medical assessments on children's overall school readiness, language skills, motor skills and socio-emotional maturity around their 6th birthday.[6] These measures serve as outcomes (measures of child development) in Chapter 3 and 4 of this thesis.

Finally, Chapter 5 of this thesis also draws upon administrative data. We combine *census data* of the years 1970 and 1987 containing the full population with data on all primary and lower secondary schools from 1964 to 1986 in the region of study (the German state Saarland). These data provide us with key figures like the numbers of male and female students and teachers, the number of classes, school type, denomination and address as well as with outcomes of interest like attained degree and employment status.

1.6 Methods

When studying effects of care and education on individual development, several sources of endogeneity problems that might lead to biased results arise. For example, reverse causality could be a problem when looking at the effect of after-school care on children's socio-emotional stability. It might be that especially children who are assessed to be socio-emotionally mature by their parents are sent to after-school care since their parents are convinced that they are mature enough to spend also the afternoon in an extra-familiar setting. Then, however, after-school care would not be responsible for children's maturity. Another problem is non-random selection into the treatment: for example, the parental decision to send a child into child care cannot be regarded as exogenous so that selection plays a role. Especially working, better educated parents might send their children into child care. These characteristics might correlate with unobservable characteristics that are important for child development

[6] Note that these medical assessments are not binding for parents. I.e. if they diagnose hyperactivity of a child, this does not lead automatically to a prescription of medicine for this child, but just recommends to take the child to an appointment at the pediatrician. This is in contrast to other studies in health economics which use prescription/treatment data of insurances as this is done, for example, in the study of Schwandt and Wuppermann (2016) on ADHS diagnoses among children.

such as parenting skills. Omitted (unobservable) variables could also play a role when evaluating effects of multigrade schools. If children living in rural areas where multigrade schools are standard have different unobserved characteristics compared to children living in urban areas where multigrade schools are not standard, estimated effects of multigrade schools might be biased. In the four contributions of my thesis, I employ several solutions to circumvent such endogeneity problems. In Chapter 2, we employ a value-added approach, i.e. a child's development measures in the current period are regressed on after-school care attendance in the previous period as well as on a set of lagged development measures. Under this approach, the dependent variable cannot influence the after-school care arrangements in the previous period. We furthermore control for a set of lagged outcome measures. This accounts for the fact that after-school center-based care attendance could be caused by a child's prior performance which is at the same time strongly correlated with a child's current performance. Lagged outcome measures are also proxies for prior observed and unobserved inputs into children's development function.

In Chapters 3 and 4, a fixed effect approach exploiting within municipality (resp. within state) variation is used. In Chapter 3, the identification of effects of the full-day care expansion relies on municipality-specific deviations from the overall trend in the timing of the expansion. In other words, we use the within-municipality variation in the supply of full-day slots over time net of the general trend in expanding the supply of full-day slots and net of any variation in the socio-economic and demographic composition of the municipalities over time to identify the effect of full-day care on child development. This approach controls for all intervening unobserved factors at the municipality level as long as these factors are constant over time. In Chapter 4, I also use a fixed effect approach to evaluate state regulations' impact on child care quantity and quality by including state and year fixed effects.

In Chapter 5, our empirical question refers to the comparison of the performance of students in a multigrade environment to a single-grade environment. The identification of the effect of a multigrade school setting relies on a triple differences (DDD) model that exploits plausibly reform-induced exogenous variation in the probability to be a multigrade student across time, region and age group.

1.7 Outlook

The studies in Chapter 2 and Chapter 3 reveal substantial effect heterogeneity of public child care. The results in Chapter 2 show that especially children from a more disadvantaged family background benefit from attending after-school care centers. In particular, children whose mothers have only a lower educational degree and children of low-income families benefit in terms of their development. Children from a more advantaged family background, in terms of maternal education and higher family income, do not exhibit any significant returns to after-school care attendance. In the same vein, the subgroup analysis in Chapter 3 reveals that children from disadvantaged backgrounds, in particular, children of immigrant ancestry, from low educated families and single parent households are affected by the expansion of full-day kindergarten, while children from more advantaged backgrounds are not affected at all, which is most likely explained by their parents being more successful in compensating for the loss of time during the afternoon. Importantly, the subgroup analysis also reveals that children of immigrant ancestry are better prepared for primary school – a finding which is most likely explained by center staff being better skilled to help immigrant children with language gaps than their parents. The effect heterogeneity of public child care is of importance to policy makers especially in light of the demographic change that predicts for most OECD countries an increase of the share of immigrant children in the education system (OECD, 2015). Given that research shows that child care attendance helps to close the immigrant-native achievement gap, public child care can be used as integration vehicle from the early years on. Another predicted demographic change is the increase of single parent households (OECD, 2011). As Chapter 3 shows, young children growing up in single parent households are more vulnerable to a reduction of time spent with the mother. These children and their parents might benefit from family counseling to cope with parenting difficulties caused by single parenthood or low income, for example. Therefore, an important future research topic will be to find out which setting can really help disadvantaged families and their children. The well-known Perry Preschool Project (Heckman *et al.*, 2010) which provided large returns for par-

ticipating children consisted of both, center-based care and family counseling. In Germany, projects trying such an integrated approach of care and family counseling (*Frühe Hilfen* (early support) and *Familienzentren* (family centers)) have been launched recently.[7] An evaluation of these projects could deliver insightful results. Another still open question is the role of group size versus the qualification of staff for the respective development of different skill dimensions (cognitive and non-cognitive skills). Would a much smaller group size enabling a closer relationship with the educational staff reduce the negative consequences of full-day care on the socio-emotional development we find in Chapter 3? The findings of Chapter 4 on the positive effect of the expansion of the family day care share – a care model with a very small group size – on socio-emotional maturity indicates that this might be the case. A follow-up project will investigate this channel in more depth. In this thesis, I do not only find heterogeneity in treatment effects regarding parental background but also regarding gender. Chapter 5 shows that girls benefited from the abolition of denominational schools and the contemporaneous disappearing of multigrade classes by more than boys did. This finding provides evidence that educational reforms do not necessarily have gender-neutral effects. Gender-specific returns to education are also a topic of further research which will evaluate a recent reform in Germany that introduced multigrade classrooms in primary school. A special focus of the analysis will be on the existence of heterogeneous effects by gender to find out whether these are already existent at a very young age.

[7] See www.fruehehilfen.de and www.familienzentrum.nrw.de for more information.

Chapter 2

After-School Center-based Care and Children's Development

This chapter is joint work with Christina Felfe from the University of St. Gallen and has been published in the B.E. Journal of Economic Policy and Analysis in 2014 (Felfe and Zierow, 2014).

2.1 Introduction

Labor market participation of women with school-aged children is nowadays common practice in most mature economies: 76% of all US American mothers with school-aged children between 6 and 18 years old are employed (Bureau of Labor Statistics, 2011); in the European Union, the employment rate of mothers with primary school-aged children between 6 and 10 years old amounts to 70% and in Germany to 75% (Eurostat, 2011). While the overall employment rate is similar, the full-time employment rate varies greatly across these regions: In the US, 58% of mothers are having a full-time job, whereas 43% do so in the European Union. In Germany, only 23% of mothers are

employed at a full-time basis. One reason often stated for not taking up a full-time job is the difficulty to coordinate work with the care facilities available for children once the school day ends (Heimer *et al.*, 2011).[1]

As a result, the main care provider for primary school-aged children is still the mother. In Germany, for instance, 64% of all primary school-aged children are looked after by their mothers.[2] Alternative care providers are relatives (26%), friends (7%) and nannies (7%). In addition to these informal arrangements, 26% of all primary school-aged children are attending after-school care centers.[3] These centers are provided by municipalities and non-profit organizations and allow mothers to participate in the labor market while their children are taken care of in a regulated setting. Yet, do children benefit or rather suffer from attending such centers in terms of their school performance and socio-behavioral development?

The answer to this question depends crucially on the quality of the child care center as well as the quality of the alternative care modes. In after-school care centers, most staff holds a pedagogical degree and supervises the children through the afternoon hours. Groups, however, are rather large with 17 to 23 children per group with 1 to 2 pedagogues. As described above, the main alternative care mode is the mother or another family member. As a consequence, the quality of the care provided might be very heterogeneous and might vary depending on the family background. For instance, a child from an advantaged family background might receive high attention and learning support as well as attend high-quality leisure activities. In contrast, less educated or less advantaged mothers might possess little knowledge about development-enhancing activities and a lower capacity to help with children's homework. Thus, the impact of center-based care attendance might vary substantially across children from different family backgrounds and is ultimately an empirical question.

[1] Table 2.6 in the appendix provides information about the length of a school day for several European countries and the US. Comparing the length of the typical primary school day between Germany and European countries as well as the US reveals that school days are relatively short in Germany.

[2] Own calculations based on the German Socio-Economic Panel, 2011.

[3] These numbers do not add up to 100% as families often rely on combinations of the alternative care modes.

The present paper provides novel empirical evidence about the relationship between children's development and after-school center-based care. Our analysis is based on data of the German Child Panel which contains information on children's school grades and measures of children's socio-behavioral development, center-based after-school care attendance, and individual background characteristics of primary school-aged children between 2002 and 2005. Our estimation targets first the mean effect of center-based after-school care on children's school performance and socio-behavioral development. Yet, to do justice to the heterogeneity in counterfactual care modes, we distinguish between different family backgrounds. In particular, we stratify with respect to maternal education and household income.

The main challenge our empirical analysis faces is non-random selection into care. First of all, the major reason for sending a child to center-based care is mothers' labor force participation.[4] There are several features that might characterize working mothers that simultaneously influence their child caring activities and thus, their children's development. For instance, higher educated or more motivated women are more likely to work. One further reason for participation in after-school center-based care might be a child's school performance. In other words, we might face reverse causality. Given such non-random selection into care, a simple mean comparison might reveal diverging performance between children attending after-school care centers and children not attending after-school care centers, while this is not necessarily the result of center-based care attendance.

To deal with these issues, we exploit the longitudinal nature of our data and employ a value added approach. In other words, we regress a child's development measures in the current period on after-school care attendance in the previous period as well as on a set of lagged development measures. This approach has the following advantages. First, it splits consecutive periods and analyzes the impact of after-school center-based care attendance in the previous period on a child's development in the current period. By default the dependent variable - alternative indicators for a child's development - cannot influence the after-school care arrangements in the previous period. Second,

[4] 80% of the mothers in our data state their labor force participation as the main reason for enrolling their children into after-school center-based care.

it controls for a set of lagged outcome measures accounting for the fact that after-school center-based care attendance might be due to a child's prior performance which in turn is a strong predictor for a child's current performance. In addition, lagged outcome measures are well-suited proxies for prior inputs, both observed and unobserved ones, into children's development function. As a result, any variables influencing children's after-school center-based care attendance and children's development up to this point should be captured by the lagged outcome measures. Yet, in case there are any further unobserved determinants of children's after-school center-based care attendance and children's current development that are neither captured by the set of lagged outcome measures nor by the included control variables - e.g. future employment plans of the mother or marital instability -, the estimates of the value added approach might still be prone to omitted variable bias.[5]

Our analysis does not reveal any significant effects of participation in after-school center-based care on children's development on average. Estimates based on subgroup analysis, however, point towards significant benefits from after-school center-based care attendance for more disadvantaged children. In particular, children of less educated mothers and of low-income families benefit from attending after-school center-based care in terms of their socio-behavioral development.

Existing research on after-school programs has focused mainly on special programs targeted at disadvantaged children in the US. Heavily evaluated programs are the "Extended Services School Initiative" - a comprehensive after-school program at 60 US-American schools targeted at high-need children in 1st to 8th grade - or "LA's Best" - an "After School Education and Enrichment Program" established in 1988 in LA's poorest elementary schools. Regarding the first program, Grossman *et al.* (2002) establish beneficial effects of the program in terms of decreased risk-taking behavior, improved school attitudes and grades, extended social networks and higher self-confidence. Brooks *et al.* (1995) analyze the "LA's Best" program during its initial phase and find that the program helped the participating children to catch up with non-attending children in terms of academic development. Huang *et al.* (2000) evaluate the intermediate and longer-run effects of the program. Conditional on

[5] For a discussion on the likelihood and direction of such a bias please refer to Section 2.4.

gender, ethnicity, English proficiency and family income, they find positive effects on children's propensity to perform well in standardized tests and children's attitudes towards school. A further study by Huang *et al.* (2008) distinguishes with respect to the intensity of program exposure (i.e. days of attendance during a school year) and finds positive correlations between exposure and students' math achievements.

Evaluations of targeted programs suggest that after-school center-based care has a positive impact on child development. Yet, based on their findings, one can hardly infer anything about the effects of universal after-school care programs. In targeted programs, the participating children come from rather disadvantaged family backgrounds and enjoy high-quality programs directly aligned with their needs. Thus, the effects of those programs are likely to present an upper bound of potential effects of universal after-school care (Blau and Currie, 2006).

To the best of our knowledge the only study that estimates the effect of non-targeted after-school supervision on child outcomes is the one by Aizer (2004). Her paper, however, focuses on the consequences of unspecified adult supervision for the behavioral development of children aged 10 to 14 years. The chosen empirical strategy - family fixed effects - allows Aizer to account for non-random selection of mothers into the work force and thus non-random selection into differential arrangements of after-school care. She finds that children with adult supervision are less likely to skip school, use alcohol or marijuana, steal or hurt someone.

Our study contributes to the literature in at least four ways: First, to the best of our knowledge, we are the first to evaluate the impact of universal after-school center-based care. As pointed out above, the existing literature focuses only on targeted child care programs or on unspecified adult supervision.[6] Second, our study focuses on both school performance as well as socio-behavioral development, and thus provides a broad picture of the effects of after-school care provision on several development dimensions important for later success in life. Third, we exploit individual panel

[6] There exist several evaluations of universal center-based care for pre-school children (Baker *et al.*, 2008, Cascio, 2009, and Fitzpatrick, 2008 for North-America; Dustmann *et al.*, 2012, Felfe *et al.*, 2015 and Havnes and Mogstad, 2011 for several European countries). Yet, to the best of our knowledge, there is no study analyzing the impact of universal care for primary school-aged children.

data and employ a value added method to address the problem of endogenous selection into after-school care. Forth, we stratify our sample along several important determinants of child development. Doing so is crucial as it reveals which children benefit and which children loose from attending after-school care centers. Finally, in contrast to the existing literature which focuses exclusively on the U.S., we focus on a European country, namely Germany.

The remainder of this study is structured as follows. The following section 2.2 provides some basic information on after-school center-based care in Germany. Section 2.3 introduces the dataset and provides descriptive statistics. Section 2.4 explains the empirical strategies, while section 2.5 presents the results of the estimations. Section 2.6 concludes.

2.2 Institutional Background

After-school center-based care in Germany exists in four different forms: After-school care clubs, the so called "Hort", are the most common mode of a formal care setting for primary school-aged children. Children go there as soon as school is over, have lunch, do their homework, can play, etc. The other three concepts are lunchtime supervision, afternoon supervision and homework supervision. These types of care are often taking place in the hort or in other child care institutions under public responsibility (Riedel *et al.*, 2005). All of the four types have in common that guided supervision is provided to children after school.

Overall, there are about 455'000 slots in after-school center-based care available to school-aged children in 2002 (the year our dataset, the German Child Panel, originates from; please refer to Section 2.3 for details on the data). Given the total number of children aged 6-10 years at this time (about 3.3 million), this means that on average every seventh child has access to center-based care. Most of these formal care slots are allocated in East Germany, where 67.7% of school-aged children are offered formal care in the afternoon. Their West German peers (excluding those living in a city state)

Table 2.1: Regional Variation in After-School Center-based Care Supply, West Germany 2002

	Offer rate in % Average Average all counties	Offer rate in % County with lowest supply	Offer rate in % County with highest supply
Baden-Wurttemberg	4.8	0.6	20.5
Bavaria	7.1	0.1	32.5
Berlin (City State)	59.2	-	-
Bremen (City State)	18.3	11.6	19.9
Hamburg (City State)	24.9	-	-
Hesse	9.9	0.9	40.2
Lower Saxony	4.5	0	16.8
North-Rhine Westphalia	5.6	0.8	20.1
Rhineland-Palatinate	4.7	0.5	14.7
Saarland	6.5	3.8	9.2
Schleswig-Holstein	5.5	0.5	14.5
West Germany (without City States)	6.0	-	-
City States	34.1	-	-

Source: Riedel *et al.* (2005), Table 20 and Table 44.

face a much smaller supply: only 6% of them have the chance to get a place in after-school center-based care. This difference is striking and questions a joint analysis of East and West Germany. Given the rather small sample available for East Germany, our analysis focuses therefore on children in West Germany only.

Figure 2.1 and Table 2.1 illustrate the supply of after-school care centers in the different West German states and counties. Most states exhibit offer rates lower than 10%. Exceptions are the city states, Berlin, Bremen and Hamburg, where 1 out of 5, 1 out of 4, and even 1 out of 2 children have access to after-school care, respectively. Yet, there is substantial variation within states: in Bavaria, for instance, the minimum offer rate just exceeds zero (0.1%), while the maximum offer rate amounts to 32.5%. Such differences are mostly driven by city-countryside differences. For instance, in Munich in 2002, 1 out of 3 children has access to a slot in after-school care, while in Oberallgäu, a rather rural Bavarian county, only 1 out of 60 children has access to a slot in after-school care.

Is this low supply of after-school care met by an equally low demand for child care facilities? Employment rates of mothers living in West Germany with children aged 6-10 lie above European average (2002: 65.7%). However, what is more important

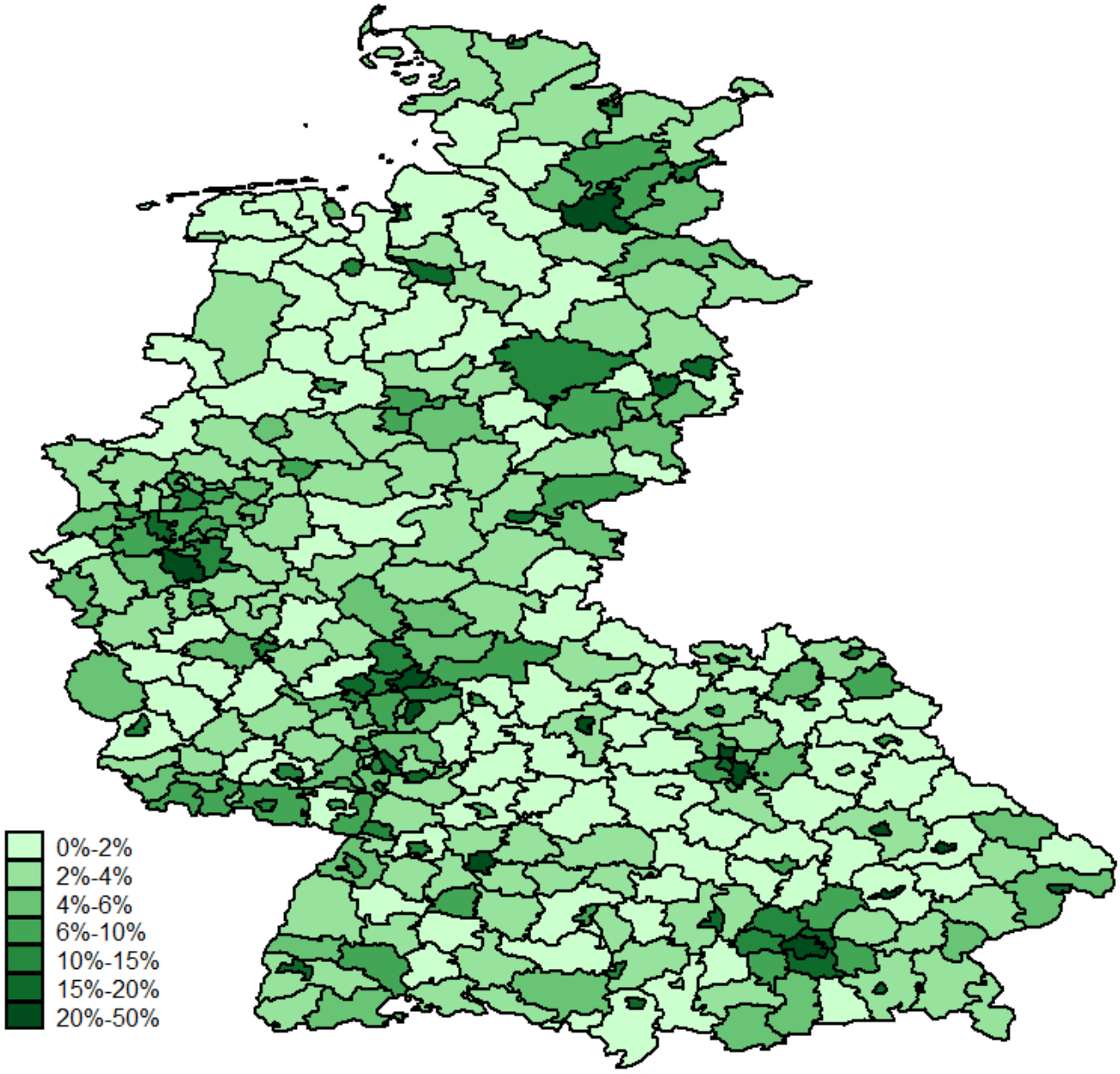


Figure 2.1: Regional Variation of After-School Center-based Care Supply in West Germany.
Note: County-specific coverage, i.e. after-school care slots as percentage of primary school-aged children.
Source: Figure based on data from Riedel *et al.* (2005).

for defining a potential demand for after-school center-based care is whether these mothers work full-time or part-time. In fact, as is shown in Table 2.2 only 16.4% of mothers in West Germany work full-time. Thus, at least 10.4% of all West German primary school-aged children have a mother who works full-time, but do not have access to a slot in center-based care (16.4% of full-time working mothers versus an after-school center-based care offer rate of only 6.0 %). In other words, we face most likely a situation of insufficient supply.[7] In response to this excess demand, public efforts are steadily rising to expand the amount of available facilities.[8]

What characterizes after-school center-based care in Germany? After-school care centers are not only a place where to keep children safe while parents are at work, but society has certain expectations about what these centers should offer. Federal law from 1990/1991 states that care centers have to promote children's development and help children to become responsible and active members of the society (§22 SGB VIII (1)). In after-school center-based care, children receive support with their homework, get lunch, do different kinds of supervised activities or simply play with their peers. The quality of supervision is relatively high. As Table 2.3 shows, on average more than two thirds of the staff hold a degree in child care, a non-university formation over three years after high school. About 6 to 8 percent of the staff even holds a university diploma in social pedagogy. Yet, in some states there is also a substantial part of staff having only completed a short-term formation in child care (less than three years), the West German average lies at 10 percent. Other measures to judge the quality of child care institutions are group size and children-per-staff ratio. The average group size in West Germany varies between 17 to 23 children per group and thus is comparable to or only slightly below the class size in primary school. The average children-per-staff ratio, however, ranges between 9 and 12.

---

[7] Further evidence for a gap between supply and demand of after-school care slots is the existence of priority setting rules after-school care centers have to adhere to when considering applications. They usually have to prioritize (1) children living in the same district, (2) children living with single mothers, (3) children with two fulltime-working parents, (4) children whose siblings are already enrolled in the center, (5) children with disabilities, and (6) children in families in distress (e.g. financial problems). Notice, that the priority criteria differ slightly depending on state and county.

[8] Table 2.7 in the appendix shows that supply increased from 1998 onwards in all West German states, with exception of Northrhine-Westfalia, where the concept of all-day-schools was established instead.

Table 2.2: Employed Mothers with Children aged 6-10 in 2002

	West Germany	East Germany
% Employed of all mothers	65.7	69.7
% Full-time employed mothers	16.4	42.3
% Part-time employed mothers	49.3	26.4

Source: Genesis-Online

These numbers make evident that the environment of after-school center-based care is probably very different from most of the potential counterfactual situations children face when not attending center-based care. First, the time in center-based care is characterized by being in a rather large group of children. This could be beneficial for children's socio-behavioral development due to peer interactions that do not necessarily take place when being at home. Yet, it could also be that the large group of children has negative consequences for a child due to a higher stress level and negative spill-over effects, in particular if children with learning or conduct difficulties are attending after-school care centers. Second, in after-school care centers there is at least one adult present who has a pedagogical background and offers educational guidance through the afternoon hours. This implies that a child attending center-based care has the opportunity to discuss her homework with a qualified pedagogue which might help the child improving in school. Given the high children-per-staff ratio and the short formation period of some staff it is, however, questionable if the time and help that can be allotted to each child is enough to "make a difference".

Care provided by mothers or family members can be very heterogenous and depends on the family background. For instance, better educated mothers may be able to provide their children with learning support as well as with development-enhancing activities - such as reading to the child, undertaking cultural activities, enrolling their children in high-quality leisure activities. Less educated mothers may not be able to provide their children with care of a comparable quality, either because they lack the ability, information or financial resources. Another dimension which might be crucial for determining inputs into children's human capital production function is

Table 2.3: Regional Variation of After-School Center-based Care Quality in West Germany

State	Group Size	Children per Staff	Pedagogical Background University degree in % of staff	Pedagogical Background Non- university formation 3 years in % of staff	Pedagogical Background Short-term formation in % of staff
Baden-Wurttemberg	22	10	8.1	61.2	3.4
Bavaria	23	10	3.7	57.1	25.8
Berlin (C)	-	-	2.5	85.4	-
Bremen (C)	20	11	17.4	65.8	3.3
Hamburg (C)	20	12	4.0	64.7	3.2
Hesse	17	10	12.6	65.3	3.5
Lower Saxony	19	10	8.4	70.8	14.9
North-Rhine Westphalia	20	9	2.7	85.3	3.9
Rhineland-Palatinate	17	9	7.2	73.7	6.7
Saarland	20	10	2.3	83.7	5.9
Schleswig-Holstein	17	10	4.6	70.0	18.3
West Germany (without C)	19	10	6.2	70.9	10.3
City States	20	12	8.0	72.0	3.3

Source: Federal Statistical Office - Kinder- und Jugendhilfestatistik.

the household income as it puts restrictions on which kind of activities a child can attend in the afternoon hours.[9]

Taken together, there might be striking differences in counterfactual care modes. As a result the impact of attendance to after-school centers on children's development may differ substantially across children from different family backgrounds. Our analysis therefore stratifies with respect to the above mentioned two dimensions: mothers' education and household income. These dimensions might serve as a good proxy for the quality of maternal or family care.

[9] In addition, the fact whether a child lives in a single-parent household or a two-parent household might be potentially crucial for the quality of care a child receives at home. Single mothers may simply lack the time to provide their children with high-quality care. Unfortunately, sample size issues (47 single mothers in our sample) prevent us from any meaningful analysis of the subsample of children who live in a single-parent household.

2.3 Data

2.3.1 Data

Our analysis is based on data from the German Child Panel (GCP). The GCP is a longitudinal survey conducted by the German Youth Institute from 2002 to 2005. Besides providing access to a wide range of information on important background characteristics, the GCP contains a broad spectrum of indicators on child development and child care provision. The survey is based on two cohorts, children born between 10/1993 and 09/1994 who were 8 years old when being interviewed for the first time, and children born between 10/1996 and 09/1997 who were 5 years old at the time of the first interview. The children and their parents were interviewed in three stages at intervals of approximately 1.5 years: 2002, 2004 and 2005.[10]

Our analysis considers interviews within the time span when children attended primary school. Therefore, we draw on the first and second interview with mothers of the older cohort, i.e. when children were between 8 and 10 years old. Following the same logic, we use the second and third interview with mothers of the younger cohort, i.e. when children were between 7 and 9 years old. Henceforth, we refer to these two interviews as the respective first and second period. We construct the sample as follows: First, we exclude all observations from children living in East Germany for reasons explained in Section 2.2. We also discard all observations in the city states Berlin and Bremen because of differences in the definition of after-school center-based care.[11] We do not keep children who are implausibly old or young in our sample.[12] Moreover, we drop all cases where information on outcomes or treatment

[10] At each stage, all mothers were interviewed and all fathers were asked to complete a questionnaire. Unfortunately, participation of fathers was very low and thus, their answers will not be considered in this paper.

[11] The reason for doing so is that the distinction between all-day school and after-school center-based care is not clear cut in those two states (after-school care often takes place inside the school building). As such, mothers' answers to the question whether their children attend after-school center-based care might not be comparable between families living in Berlin and Bremen and the remaining states.

[12] In other words, we do not consider children who deviate 2 or more years from the originally targeted age group.

are missing. Finally, we restrict our sample to children who were present in both, the respective first and second period. This leaves us with 857 children, 363 of the younger cohort and 494 of the older cohort.[13]

We create the treatment variable "afterschool" that indicates whether the child attends after-school center-based care. This variable takes the value 1 if a child uses at least one of the following after-school center-based care settings: hort, homework supervision, lunchtime supervision and/or afternoon supervision.

The GCP survey contains questions covering a broad spectrum of child development dimensions, inspired in particular by the Child Behavior Checklist CBCL (Aschenbach and Edelbrock, 1981) and the Temperament Survey by (Windle and Lerner, 1986; Wahl, 2008)). We group the numerous (14) questions according to the logic of the Strengths and Difficulties Questionnaire (SDQ), a brief behavioral screening questionnaire widely used in youth psychology research. This makes our development indicators more comparable to a recently emerging literature on the impact of early child care on children's non-cognitive development (see for instance Datta-Gupta and Simonsen, 2010). The SDQ score is built on the following four development dimensions [14]:

- *Emotions*: information on how often a child is happy, nervous, worried, easily scared, etc.
- *Conduct*: information on how often a child has temper tantrums, fights with other children or bullies them, etc.

[13] Table 2.8 in the appendix provides an overview of our sample construction. Notice that there are significant differences between the original dataset and the sample used in this study regarding (1) the household income, i.e. children dropped from the sample live more often in a household with a net income below 2250 Euro/month (12% difference), (2) the immigrant background, i.e. children dropped from the sample have more often an immigrant background (6% difference), (3) the education of the mother, i.e. children dropped from the sample have more often a less educated mother (5% difference), (4) the birth weight of the child, i.e. children dropped from the sample had more often a low birth weight (4% difference). As a result, the remaining sample might not be representative anymore. Nevertheless, if our hypothesis is true and after-school care should be especially beneficial for children with a less advantaged background, the arising selection leads to estimates that present a lower bound of the effect of after-school center-based care attendance.

[14] Table 2.9 shows the survey questions concerning children's development and how we group them into the different developmental categories.

- *Hyperactivity*: information on whether a child gets easily distracted, acts before thinking, cannot stay still for long, etc.
- *Peer Relations*: information on whether a child is generally liked by other children, is rather solitary, is bullied by other children, etc.

Following the practice of the SDQ questionnaire we construct an average score based on the four development dimensions - emotions, conduct, hyperactivity, and peer relations - and thus, obtain a measure for children's socio-behavioral development. In addition, we average a child's performance in the core subjects mathematics, reading and writing, which serves as our overall measure of school performance. We standardize all development indicators to have a zero mean and a standard deviation of one. Interpretation of the indicators is straightforward: the higher the value of an indicator the less difficulties a child has in the respective development dimension, i.e. the better is her performance in the respective dimension.

One potential disadvantage of the GCP dataset is that all developmental indicators are derived from a mother's judgment of her child's abilities and characteristics. A problem could arise if mothers sending their child to after-school care centers justify their absence and thus overestimate their children's development. Yet, such reporting problems may be primarily present for more subjectively measured outcomes, such as personality traits, and less for more objectively assessed outcomes, such as school performance. Thus, if there are only differences observable with respect to subjectively measured outcomes, but not with respect to more objectively reported outcomes, misreporting may be an issue.

A further problem might arise if maternal judgment capacity changes when children access after-school care. In other words, some developmental dimensions (like fighting with other children) may only become apparent once the child enters formal care and spends time with a larger peer group. Yet, basically all children have attended Kindergarten from age 3 onwards and thus have been exposed for a longer time to institutionalized care. Moreover, children are already enrolled in school. Thus, at this

point in time parents should have already had the opportunity to observe their child both in the family as well as in the outside-family context.

Finally, one further disadvantage of the GCP data is that children's school performance is not measured by standardized tests. The indices are based on recent school grades (which are the result of non-standardized assessments of written and oral performance in the class room). The school performance measure may thus be biased by different grading policies across the German states. Taking this into account, we control for state fixed effects in our regressions. Doing so helps us furthermore to tackle any differences in the school infrastructure and curricula as well as in the public funding for day care institutions at the state level (notice that the educational system is under the jurisdiction of the states).

Besides public care settings, there are other important determinants of children's development. The most relevant ones are a family's socio-economic status, parental education, and a child's initial endowments (Blau, 1999; Case *et al.*, 2002; Currie, 2009; Almond and Currie, 2011; Black *et al.*, 2005). Our set of control variables includes therefore gender, immigrant background, birth weight, prevalence of any kind of health disorder, pre-school attendance, mother's age, single parenthood, education of the mother, and finally household net income and number of siblings. To do justice to the literature which puts forward the socio-economic environment as a further determinant of child development (among others Kling *et al.*, 2007), we additionally consider a set of regional characteristics measured at the county level (population density, unemployment rate, female employment rate, net migration, GDP per capita and primary school-aged children as percentage of the total population)[15]. In addition, we distinguish between rather urban and rather rural counties within each state by including a dummy indicating whether the county of residence is rather urban as well as the interactions between this dummy and the set of state dummies.

[15] The regional data are from the INKAR database provided by The Federal Institute for Research on Building, Urban Affairs and Spatial Development and merged with the GCP data via county codes.

2.3.2 Descriptives

This section provides first evidence on the raw differences between *Treated* children - those who attend after-school center-based care in the first wave - and *Control* children - those who do not attend after-school center-based care in the first wave. Since the available data do not allow us to derive precise statements about the afternoon activities of control children compared to treated children, we need to be very general when describing the difference between the two groups: Children in the treatment group receive schooling in the morning and after-school center-based care in the afternoon. Children in the control group receive schooling in the morning and anything else than after-school center-based care in the afternoon.[16]

Table 2.4, Panel A displays descriptive statistics on children's development. The first column displays the means of the overall sample, column 2 and 3 display the mean values of the treated and the control group, respectively, and column 4 displays the difference between the two groups along with the respective standard error. Treated children perform on average significantly worse in school (0.17 standard deviation, henceforth sd). Treated children also display a significantly worse performance in the non-cognitive dimension "SDQ Score" (0.31 sd). Table 2.11 in the appendix shows the differences between treated and control children regarding the four components of the SDQ score. Treated children perform significantly worse regarding their emotional stability (0.18 sd), conduct (0.20 sd) and hyperactivity (0.28 sd). The difference between treated and control children's peer relations is not significant.

Table 2.4, Panel B contains information on several child characteristics. It shows that treated and control children do not differ significantly with respect to age, birth weight, health disorder and gender. Yet, foreign children and children who were already enrolled in formal care under the age of three years are more likely to be enrolled in an after-school care center (by 5 and 7.6 percentage points, respectively). Do treated and control children differ with respect to their family background? Table

[16] Table 2.10 in the appendix hints at differences regarding afternoon activities: treated children are by 8 percentage points less likely to receive private music lessons or sport training than control children. Moreover, treated children are by 9 percentage points less likely to have lunch together with their family than control children.

Table 2.4: Descriptive Statistics according to After-School Care Usage

	Pooled	**No Afterschool**	**Afterschool**	**Afterschool vs. No Afterschool**	
	N=857	N=689	N=168	Diff.	Std.Error
Panel A. Child Outcomes					
School Performance					
Grades	0.093	0.127	-0.045	-0.172**	(0.086)
Socio-Behavioral Development					
SDQ Score	-0.040	0.022	-0.292	-0.314***	(0.086)
Panel B. Child Characteristics					
Age	7.852	7.862	7.810	-0.053	(0.056)
Low Birth Weight	0.113	0.107	0.137	0.030	(0.027)
Disorder	0.183	0.174	0.220	0.046	(0.033)
Female	0.494	0.501	0.464	-0.036	(0.043)
German	0.914	0.923	0.875	-0.048**	(0.024)
Early Care Attendance	0.177	0.163	0.238	0.076**	(0.033)
Panel C. Family Characteristics					
Mother's education: primary	0.20	0.21	0.16	-0.054	(0.035)
Mother's education: secondary	0.59	0.60	0.52	-0.079*	(0.042)
Mother's education: university	0.19	0.16	0.29	0.125***	(0.033)
Mother is working	0.63	0.60	0.77	0.173***	(0.041)
Mother is single	0.05	0.03	0.14	0.109***	(0.019)
Mother's age: under 30	0.06	0.06	0.05	-0.015	(0.020)
Mother's age: 30 to 34	0.17	0.17	0.18	0.010	(0.032)
Mother's age: 35 to 39	0.40	0.41	0.39	-0.015	(0.042)
Mother's age: 40 to 44	0.29	0.28	0.30	0.021	(0.039)
Mother's age: 45 to 49	0.07	0.07	0.06	-0.013	(0.022)
Mother's age: $>$ 50	0.01	0.01	0.01	0.006	(0.007)
Child has siblings	0.83	0.85	0.77	-0.083***	(0.032)
Persons in the household	4.23	4.29	3.97	-0.319***	(0.097)
Monthly net income () 0- 1250	0.05	0.05	0.06	0.012	(0.019)
Monthly net income () 1250- 2250	0.28	0.27	0.33	0.065*	(0.039)
Monthly net income () 2250-3250	0.35	0.36	0.27	-0.091**	(0.041)
Monthly net income () $>$ 3250	0.25	0.25	0.26	0.011	(0.038)
Panel D. Regional Characteristics					
Urban county	0.26	0.23	0.37	0.141***	(0.038)
Population Density	328	289	490	201***	(59.24)
Unemployment Rate	8.84	8.73	9.30	0.568***	(0.219)
Female Employment Rate	7.98	7.92	8.25	0.327*	(0.172)
Net Migration	3.29	3.28	3.32	0.045	(0.267)
After-school care slots per child	0.07	0.07	0.11	0.039***	(0.006)
GDP p.c. in 1000euro	28.65	27.90	31.72	3.821***	(0.985)
% of Children aged 6-10	4.23	4.27	4.06	-0.210***	(0.044)

Note: The statistics are based on the first period of the GCP survey. Treated children are those children who attended after-school center-based care in the first period. Control children are those children who did not. Standard errors in parentheses.
* p$<$0.10, ** p$<$0.05, *** p$<$ 0.01.

2.4, Panel C shows that - as already highlighted before - mothers' labor force participation appears to be one of the main correlates with children's participation in after-school center-based care. We observe a significant difference in the labor force participation of mothers whose children attend and do not attend after-school care (77% versus 60%). Among those mothers who are working the ones who use after-school care for their child have significantly more often a fulltime-job (23% versus 13%, see table 2.12 in the appendix).[17] Major differences are also observed with respect to mothers' education: Mothers of treated children are less likely to possess only a primary or secondary school degree (by 5.4 percentage points and 7.9 percentage points, respectively), but more likely to have attended university (by 12.5 percentage points). There is, moreover, a large difference regarding single parenthood: 14% of mothers belonging to the treated group are single mothers. This percentage is much smaller for the control group (3%). Moreover, treated children have less often siblings than control children (77% vs. 85%), and thus also a lower number of persons living in the same household (4.3 vs. 4 persons on average). We also observe that families with a higher net income per month are less likely to send their child to an after-school care center: 32% of control children live in a family with a household income below 2250 Euro per month, while 39% of treated children do so. In contrast, 61% of control children live in a family with a household income above 2250 Euro per month, while only 53% of treated children do so. Regarding the regions of residence, children in formal care live more often in regions with a higher population density, a higher unemployment rate, a higher GDP per capita, a lower percentage of primary school-aged children, and a higher supply of after-school care slots per child. Finally, after-school care attendance is generally higher in urban areas (38% vs. 23%). As such, we control in our analysis not only for a set of state dummies, but also distinguish within states between more and less urban counties.

Overall, the above findings indicate that children attending after-school center-based care - compared to children who are not attending - stem from a more advantaged background in terms of mother's education. However, treated children have also

[17] Note that fathers of treated children are by 5.6 percentage points more likely to be unemployed. This goes in line with the priority setting policy of after-school care centers to facilitate attendance to children of families in financial distress, as described in Section 2.4.

fewer siblings and reside more often in a single-parent household, both factors that may imply that treated children are less likely to encounter social interactions at home. Moreover the parental income of treated children is lower which might influence the family's capacity to arrange development stimulating activities for their child. Thus, based on the descriptive evidence, it is not obvious whether the selection into after-school care would lead to over- or understating the impact of after-school care attendance when just comparing the raw differences between control and treated children as shown in Table 2.4, Panel A.

To understand a bit better whether families of treated children provide on average an adverse environment for a child's development, table 2.13 in the appendix compares both groups with respect to the prevalence of family problems, conflict behavior and the external help parents receive regarding their children's education. The comparison shows that mothers of treated children worry significantly more often about problems concerning housework and household duties, face more stress in general as well as related to their job and their partner - all factors going in line with the higher engagement in paid work, the lower household income and the higher prevalence of single motherhood among mothers of treated children. Yet, there are no significant differences with respect to worries about children's health or school performance. Also regarding the conflict behavior mothers of treated children do not differ significantly from mothers of control children. Finally, families do not differ with respect to the external help they receive from pedagogues or psychologists.

2.4 Econometric Framework

This section clarifies what we mean by the effect of after-school center-based care on children's development, introduces the empirical strategy and discusses the respective underlying assumptions.

We define the effect of after-school center-based care on a child's development as the difference between a child's development when the child attends after-school center-

based care and the development the same child would reach if she did not attend after-school center-based care. Notice that the counterfactual situation comprises alternative scenarios, but refers most likely to care provided by the mother or any other family member (see Sections 2.1 and 2.2 for a discussion). While it is crucial to understand the counterfactual situation, it is beyond the scope of this paper to investigate underlying mechanisms. This paper evaluates the average effect of after-school center-based care, as well as the effect for several subgroups that might experience counterfactual care situations that differ substantially in their quality.

When identifying the effect of after-school center-based care on children's development, we face several empirical challenges. First, selection into after-school center-based care is a non-random decision. One of the main reasons for sending a child to after-school center-based care is a mother's decision to participate in the labor market. In fact, 80% of mothers in our data state their employment status as the decisive reason to send their children to center-based care. If a mother's decision to work is correlated with her child caring quality or with further individual or family characteristics that directly influence children's development, the unconditional correlations between after-school center-based care attendance and children's development measures as shown in Table 2.4, Panel A do not reflect a causal effect. The direction of the arising bias depends on the relative quality of the care provided by the mother. If mostly high-skilled mothers decide to work and it is those mothers who provide their children with best care and a stimulating environment, our estimates are likely to be upward biased. Yet, if rather women from disadvantaged backgrounds need to engage in the labor market and this type of women cannot guarantee their children high-quality care and a stimulating environment, the raw correlations would underestimate the true beneficial effects of after-school center-based care. In addition, a child's prior performance might influence parents' decision to send the child to after-school center-based care. On the one hand, mothers might opt out of work in case their child does not perform sufficiently well. On the other hand, insufficient school performance or conduct problems might constitute a reason for parents to actually send their child to after-school center-based care. A simple "selection-on-observables" approach can control for all observable characteristics that determine

after-school center-based care attendance and children's development simultaneously. However, the estimates of this approach would not give us the true effect if there are unobservable characteristics that impact both the decision to send a child to after-school center-based care attendance and a child's development.

The longitudinal nature of the data used in this study allows for two alternative approaches that can handle at least partially the problems arising due to non-random selection into after-school center-based care based on unobservable characteristics. The first approach is the well-known fixed effect framework (FE). The advantage of the FE method is that it eliminates any time-invariant unobservable characteristics that might confound with children's development and after-school center-based care attendance. The disadvantage of the FE framework is that it fails to deal with i) reverse causality and ii) endogeneity of after-school center-based care attendance with respect to previously acquired skills. Regarding the first issue: Besides maternal labor force participation, a child's development constitutes a decisive reason for sending a child to after-school center-based care. Hence, reverse causality is likely to exist in our setting, i.e. a child's school performance as well as behavior causes her after-school center-based care attendance and not vice versa. Regarding the second issue: According to the human capital production theory (Cunha *et al.*, 2006), skill formation follows a dynamic process, i.e. earlier acquired skills boost the formation of later skills. As a result, earlier skills might not only determine after-school center-based care attendance, but also the formation of current skills and thus are likely to be confounding variables. A FE framework exploits the changes in outcome and control variables, but fails to model the impact of the level of the lagged outcome measures on the current level of the outcome variable. This dynamic can be captured by the second approach - also called the value added approach (Todd and Wolpin, 2007). This approach regresses children's development measures in the current period on after-school care attendance in the previous period as well as on lagged development measures. This procedure has two main advantages: First, by focusing on after-school care attendance in the previous period it rules out reverse causality. By definition, current development cannot influence previous after-school center-based care attendance. Second, by including the set of lagged development indicators as control

variables the model respects the concepts of the human capital production function - children's current development is the outcome of previously acquired skills - and the fact that after-school center-based care attendance is likely to depend on children's prior development. Thus, this approach helps us to tackle not only any bias arising due to selection into after-school care based on prior performance, but also any bias arising due to unobserved input factors that enter into a child's development up to the previous period. The latter is due to the lagged outcome measures capturing any inputs into children's development up to the first wave. These inputs can be both observed - such as parents' education or household composition - and unobserved - such as parental attention or affection towards the child - and are likely to depend on the actual or perceived personality of the child, e.g. her innate ability, her behavior, her physical or psychological robustness.

Given the described advantages and disadvantages of the alternative methods based on longitudinal data, we believe that the value added approach fits better the problems arising when modelling children's development process. The equation to be estimated can thus be expressed as follows:

$$\begin{aligned} y^n_{ict} = \alpha_0 + afterschool_{it-1}\delta + \sum_{n=1}^{N} y^n_{it-1}\theta_n + X_{it-1}\alpha_1 + X_{ct-1}\alpha_2 \\ + \sum_{s=1}^{9} S^s_{t-1}\alpha_3 + \sum_{s=1}^{9} S^s_{t-1} * urban_{ct-1}\alpha_4 + e_{ict} \end{aligned} \tag{2.1}$$

In this equation y^n_{ict} represents child i's n^{th} development outcome at time t in county c. $afterschool_{it-1}$ is the treatment variable and indicates whether the child at time t-1 attends after-school center-based care or not. X_{it-1} represents child and family characteristics at time t-1, while X_{ct-1} represents regional characteristics (at the county level) at time t-1. Regarding child characteristics we control for a child's age, gender, low birth weight, innate disorder, nationality, early care attendance, and number of siblings; regarding mother characteristics we control for her age, education and single parenthood; regarding family characteristics we consider net household income and

regarding county characteristics we consider population density, GDP per capita, net migration, unemplyoment rate, female employment rate, and number of primary school-aged children in each county. Finally, we control for a set of state dummies S^s_{t-1} and distinguish in each state between more and less urban counties (by including the set of interaction terms between urban counties and state dummies). Notice, that to circumvent any bias due to endogenous controls (Angrist and Pischke, 2008) all individual and regional control variables are also taken from the previous period.

The coefficient δ measures the impact of after-school center-based care on child development as long as the following assumptions are fulfilled. First of all, there is no immediate impact of after-school center-based care attendance on children's development: children's development in the previous period is assumed to be unaffected by center-based care attendance in the previous period. In other words, the effect of attending after-school care only becomes apparent after a certain time of attendance. This assumption might seem quite demanding, in particular in terms of children's socio-behavioral development (school grades might indeed refer to performance in the last exam or even the grade received in the last school year). Nevertheless, if this assumption does not hold true and there is an immediate impact of after-school center-based care on a child's development, coefficient δ in equation 2.1 provides us with a lower bound for the effect of after-school center-based care attendance as part of the effect is then already captured by the coefficients of the lagged outcome. Second, there shall be no further unobserved variables that determine center-based care attendance in the previous period and children's development in the following period that are not captured by the set of development measures in the previous period and the included control variables. For instance, any change in a child's life that might be foreseen by the parents, but not yet experienced by the child - e.g. future labor force participation of the mother, arising marital instability -, might constitute such a determinant of parents' decision to send their child to center-based care and

influences their child's later development. This scenario requires, however, perfect foresight as well as capability of parents to hide a future change to the child.[18]

Finally, as discussed above, the impact of child care is likely to vary with the counterfactual care situation, i.e. with the quality of care children would receive when not attending center-based care during the afternoon hours. Unfortunately the quality of care is not or at best imperfectly observable for the econometrician. Yet, to address at least partially existing effect heterogeneity, we stratify our analysis according to several dimensions that are not only major determinants of children's after-school center-based care attendance (see Section 2.3), but also of the quality of the counterfactual care mode (Davis-Kean, 2005; Linver *et al.*, 2002; Yeung *et al.*, 2002). To be more precise, we distinguish with respect to mothers' education - primary education, secondary education and university education. Typical occupations one can work in when holding a primary education degree is a cashier or a hairdresser, when holding a secondary degree a nurse, a child carer or physiotherapist, and when holding a university degree a school teacher, a physician or an engineer. In addition we distinguish with respect to families' net monthly household income - a monthly net income below 2250 Euros, between 2250 and 3250 Euros, and more than 3250 Euros.[19]

2.5 Results

How does participation in after-school center-based care influence children's development? This section presents the results of the value added approach shown in equation 2.1. We show estimates for the pooled sample as well as for two alternative stratifications: we stratify with respect to (1) the education of the mother, i.e. primary, secondary or university degree; (2) the household income, i.e. net monthly household

[18] In a previous version of this paper we also employ an instrumental variable strategy using the supply of after-school center-based care at the county of residence as an instrument. Unfortunately, regional variation in after-school center-based care does not constitute a sufficiently strong enough instrument to provide precise estimates. We therefore do not present the results based on the instrumental variable estimations in the paper, but provide them upon request.

[19] The three income groups are not chosen adhoc, but correspond to the income brackets underlying the question asked to the parents participating in the GCP.

income of the child's family below 2250 Euro, between 2250 and 3250 Euros or above 3250 euro.

2.5.1 Baseline Results

Table 2.5 reports the results based on the value added approach regressing school performance and socio-behavioral development on previous after-school center-based care attendance, on all previous development dimensions (school performance and socio-behavioral development), on previous individual and regional characteristics as well as on a set of state dummies and interactions between each state dummy and a dummy indicating whether the county of residence is urban. Column 1 shows the results with respect to children's school performance (grades) being the dependent variable, column 2 shows the results with respect to children's socio-behavioral development being the dependent variable.

Could it be that mainly previously worse performing children are sent to after-school care centers? Or is it the case that mainly children from advantaged backgrounds are sent to after-school center-based care? Under any of these scenarios, the raw differences shown in Table 3.1 are misleading. The estimates based on the value added approach are purged - at least in terms of observable characteristics - from any bias due to selection into after-school center-based care. As we can see in Panel A of Table 2.5, the average child attending center-based care does not perform significantly worse neither with respect to her school performance nor to her socio-behavioral development. In terms of correlations between children's development and further background characteristics we observe the following (see Table 2.14 in the Appendix): Girls exhibit generally better school grades (by 0.17 sd) as well as socio-behavioral development (by 0.24 sd). As expected, mothers' education is highly correlated with children's cognitive development: in comparison to children whose mother has a secondary school degree, children whose mother only has a primary educational degree perform significantly worse in school (by 0.19 sd), while children whose mother has a university degree fare better (by 0.10 sd). Children's school performance is moreover

Table 2.5: Value Added Approach Results

	School Performance	*Socio-behavioral Development*
Panel A: Pooled Sample	0.064	0.022
Previous After-School Care Attendance	(0.076)	(0.075)
N=857 (168 treated, 689 control)		
Panel B: Education of Mother		
Primary Education	0.319	0.448**
Previous After-School Care Attendance	(0.235)	(0.226)
N=175 (27 treated, 148 control)		
Secondary Education	0.104	0.145
Previous After-School Care Attendance	(0.101)	(0.099)
N=503 (88 treated, 415 control)		
Uni Degree	-0.227	-0.210
Previous After-School Care Attendance	(0.164)	(0.183)
N=156 (47 treated, 109 control)		
Panel C: Income of Family		
Low Income	0.190	0.261**
Previous After-School Care Attendance	(0.125)	(0.121)
N=284 (66 treated, 218 control)		
Middle Income	0.168	-0.012
Previous After-School Care Attendance	(0.151)	(0.155)
N=297 (46 treated, 251 control)		
High Income	-0.108	-0.154
Previous After-School Care Attendance	(0.158)	(0.157)
N=217 (44 treated, 173 control)		

Notes: This table summarizes the results of the value added regressions. Panel A shows estimates of the after-school care coefficient using the whole sample of 857 children. The first column shows the estimate for the regression of school performance in the second period on after-school care attendance in the first period, both development measures in the first period, individual and regional characteristics in the first period and a set of state dummies and interactions between each state dummy and a dummy indicating whether the county of residence is urban. Individual characteristics include characteristics of the child (age, gender, low birth weight, innate disorder, nationality, and early care attendance), the mother (age, education and single parenthood), and the family (number of siblings and net household income). Regional characteristics refer to population density, GDP per capita, net migration, unemployment, female employment, and the number of primary school-aged children in each county. The second column shows the result of a corresponding regression with socio-behavioral development as the dependent variable. Panel B shows the respective regression results stratified by mothers' education. Panel C stratifies with respect to the net monthly household income (categories that are used in the survey: below 2250 euro, between 2250-3250 euro, above 3250 euro). Standard errors are in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$
Source: Own calculations, based on the German Child Panel.

monotonically increasing in household's net income: children whose families receive between 2250 and 3250 Euros net per month fare significantly better than children in families with an income lower than 1250 Euros net per month (by 0.26 sd); children whose families receive more than 3250 Euros per month exhibit the best grades (by 0.31 sd higher than the performance of children from the poorest families).

Do the estimated correlations between lagged and current development measures support the predictions of the human capital production theory? Do existing skills boost the development in the same skill dimensions or even in further skill dimensions? In terms of children's cognitive performance we observe the following: Grades are significantly correlated with previous grades (0.52), not, however, with previous socio-behavioral outcomes. Interestingly, previous school performance is significantly related with children's socio-behavioral development (0.05 sd), yet to a much lower extent than the inter-temporal correlation between previous socio-behavioral development and current socio-behavioral development (0.50 sd). Thus, the estimated correlations between previous skills and current skills provide some evidence in favor of the predictions of the human capital production theory.

The estimates discussed so far display average effects for the impact of center-based care attendance on children's development. Yet, to which extent do these average effects hide differential effects for subgroups? Does after-school center-based care constitute for certain children a more beneficial care mode than the counterfactual care provided by their mother? Panel B and Panel C in Table 2.5 shed some light on potential effect heterogeneities across the following subgroups: mothers with primary, secondary and university degree (Panel B), as well as low, intermediate and high household income (Panel C). The characteristics underlying these strata did not only prove to be major determinants of children's attendance to after-school care centers (see Section 2.3), but also represent family backgrounds that are likely to differ in the quality of care the mother or any other family member may be able to offer children.

The estimates in Panel B in Table 2.5 show that children of the least educated mothers exhibit significant and positive returns to after-school care attendance in terms of their socio-behavioral development (0.45 sd, significant at the 5 % significance level). In

contrast, after-school care centers do not seem to constitute a superior care mode for children of mothers with a higher educational degree. A similar gradient in returns to after-school center-based care is observed when stratifying with respect to family income: the higher the family income the lower the gains from after-school center-based care attendance (see Panel C of Table 2.5). Children of the poorest families gain the most when attending after-school center-based care: the gains in terms of socio-behavioral development amount to 0.26 sd (significant at the 5 % significance level). Children of wealthier households, however, do not gain from after-school center-based care attendance: none of the coefficients with respect to socio-behavioral development or school performance is significantly different from zero.

The results for these stratifications provide some supportive evidence for the relevance of the quality of the counterfactual care mode. Highly-educated mothers and wealthier households seem to be able to provide their children with a more stimulating supervision and program during the afternoon hours and thus, attending after-school care centers seems to be less beneficial for their children. In contrast, less educated or less wealthy mothers seem to lack the capacities or opportunities to provide their children with learning support or further stimulating activities, and thus their children indeed seem to be better off when attending after-school care centers.

Before discussing the policy relevance of our results, we would like to point out that our results are robust to controlling for the whole set of county dummies instead of controlling for the set of state dummies, the respective interactions with the dummy indicating urban areas and all regional characteristics of the county of residence (see table 2.15 in the Appendix). Also under this specification, the average child does neither display significant gain nor harm from after-school center-based care. Yet, stratifying with respect to mothers' education and household net income reveals again a socio-economic gradient in the returns to after-school center-based care attendance: children from the lowest socio-economic background have the most to gain when being placed in after-school center-based care, while children from the best socio-

economic background have the most to loose when being placed in after-school center-based care.[20]

2.5.2 Discussion

The empirical analysis conducted in this study reveals two important findings. The first important finding of our study is that on average after-school center-based care attendance does neither benefit nor harm children's school performance and socio-behavioral development. This finding stands in stark contrast to the first impression gained from a raw mean comparison between the development of children attending after-school center-based care and children not attending after-school center-based care. Yet, drawing the conclusion that investing money into expanding after-school center-based care is like spending money for nothing would be a premature conclusion. Indeed, the second and possibly more interesting finding of our study is that only some subgroups benefit from after-school center-based care. To be more precise, only children from less advantaged families benefit from attending after-school center-based care, while children from more advantaged families are unaffected by attending after-school center-based care.

Yet, can after-school center-based care help to overcome existing disparities between children from different family backgrounds? In our sample the socio-economic gap[21] with respect to children's school performance and socio-behavioral development amounts to 0.3-0.4 sd. Based on our results, children from the lowest socio-economic background gain between 0.26 and 0.45 sd in terms of their socio-behavioral development when attending after-school center-based care. Gains in terms of school performance are also positive and range between 0.19 and 0.32 sd, but are not significantly different

[20] All together there are 76 counties represented in our sample. As such, the latter specification considers 97 control variables (76 county dummies and 23 control variables already included in the baseline specification). Given the rather small sample size, in particular in some of our strata, we abstain from using the latter specification as our baseline specification.

[21] The socio-economic gap refers to the difference between children from the lowest socio-economic background (e.g. mothers with primary education only and families with a net income of less than 2250 Euros per month) and the highest socio-economic background (e.g. mothers with university education and families with a net income of more than 3250 Euros per month)

from zero. In other words, our results indicate that after-school center-based care attendance helps to close the socio-economic gap at least in terms of socio-behavioral development.

Taken together, our results provide crucial insights for a design of further expansions of after-school care: expanding after-school care can contribute positively to the development of children, in particular to the development of children of disadvantaged groups. If the target is to overcome existing disparities between children from different family backgrounds a clear priority setting for accepting children into after-school center-based care is necessary. Such a priority setting should be based on criteria such as mothers' education or household net income, proxies for the opportunities and quality of care families might be able to offer to their children. Yet, if the policy aim is to design after-school care centers that guarantee positive returns for all children, independently of their family background, the quality of after-school care centers should be the main objective of future policy reforms. In particular, the quality of care provided by after-school care centers has to match if not even to exceed the quality of care provided by mothers from the best socio-economic families. Nevertheless, to provide concrete policy recommendations further research and in particular research on the counterfactual care modes is necessary.

2.6 Conclusions

How does attendance to after-school care centers affect children's development? In light of increasing employment rates among women with primary school-aged children this question is highly policy relevant.

The present study sheds some light on this question using data from the German Child Panel. We exploit the longitudinal nature of this data and employ a value added approach to address the problems of reversed causality and endogenous selection into after-school center-based care. Our findings do not point to any significant effects of after-school center-based care attendance on the average child. Subgroup analysis,

however, reveals substantial effect heterogeneity. Children from a more disadvantaged family background benefit from attending after-school care centers. In particular, children whose mothers possess only a lower educational degree and children living in a low-income family benefit in terms of their socio-behavioral development. On the contrary, children from a more advantaged family background, in terms of maternal education and higher family income, do not exhibit any significant returns. The results presented in this study provide useful insights for current policies aimed at expanding the supply of after-school care centers. First, the priority setting for acceptance to after-school care centers should be clear: priority should be given to children who if not attending after-school care centers would be exposed to a less stimulating environment. Those are children from the most disadvantaged families (in terms of maternal education and family income). Such a priority setting may actually help leveling the playing field between children from the least and most advantaged families. Second, the quality of care provided in after-school care centers should be at the heart of the discussion about expanding after-school center-based care. Only if the care provided in after-school care centers matches or even exceeds the care offered by the most advantaged families, one can make sure that every child benefits from publicly provided after-school center-based care.

Appendix

Table 2.6: Length of Primary School Days in Europe and in the US

Country	Length of the school day	Lunch	Recent developments
France	8.30am-4.30pm	Lunch in school	
Germany	7.30am or 8.30am-12.30pm or 1.30pm	Lunch only if all-day school	In 2003 4 billion program to foster all-day opening of schools, i.e. until 4pm or 5pm at least 3 times per week. Yet, the expansion of all-day schools is very slow.
England	9am-4pm,	One hour break at lunchtime	In 2002 education act: children can come earlier (8am) or stay longer and enjoy extra activities (6pm)
Spain	9am or 10am-12am, then 2,5h break, then 2pm or 3pm-4pm or 5pm	Lunch in school possible	Non-stop day approach: 9am-2pm
Italy	Only in the morning if 24h per week or morning and afternoon if more hours per week	Lunch in school if more than 24h per week	Since 2009/10: Parents can decide: only mornings or additional afternoon classes
Netherlands	8.30am-3pm or 4pm, only to 2pm if non-stop day	One hour break at lunchtime	Primary schools increasingly opt for non-stop day, then lunch in class
USA	8.30am-3pm	One hour break at lunchtime	

Source: Kamette (2011) and Patall *et al.* (2010)

Table 2.7: Supply of After-School Care Slots Across States and Time

State	Offer rate 1998	Offer rate 2002	Offer rate 2007	Change 1998 to 2002 in %	Change 2002 to 2007 in %
Baden-Wurttemberg	3.2	4.8	5.4	48.2	12.9
Bavaria	5.7	7.1	10.6	24.7	48.6
Berlin (C)	-	59.2	-	-	-
Bremen (C)	-	18.3	18.0	-	-1.7
Hamburg (C)	23.0	24.9	28.0	8.2	12.5
Hesse	8.3	9.9	12.4	19.6	25.2
Lower Saxony	3.5	4.5	5.1	29.3	14.0
North-Rhine Westphalia	4.5	5.6	4.5	24.7	-20.7
Rhineland-Palatinate	3.3	4.7	5.4	42.9	12.9
Saarland	3.5	6.5	7.6	88.6	16.8
Schleswig-Holstein	4.8	5.5	7.0	15.8	26.5
West Germany (without C)	4.6	6	7.2	33.6	14.7
City States	-	34.1	-	-	-

Source: Riedel *et al.* (2005), Table 20 and Table 44, and Federal Statistical Office: Kinder- und Jugendhilfestatistik

Table 2.8: Sample Construction

	Young Cohort Children born 1996/1997	Old Cohort Children born 1993/1994
Original Sample, first period*	1148	1042
Without East Germany, Berlin and Bremen	939	833
Interview in second period** took place	545	575
Child is not implausibly old or young	544	568
Outcome and treatment data are available in first period	377	520
Outcome and treatment data are available in second period	363	494
Final Number of Observations	**363**	**494**

Note: This table describes the construction of the final sample we use for our analysis.
*i.e. second wave of young cohort (2004) and first wave of old cohort (2002)
**i.e. second and third wave of young cohort (2005) and first and second wave of old cohort (2004)

Table 2.9: Questions about Children's Development Used in the GCP Survey

Emotional Skills:
Is your child sometimes sad?
Is your child sometimes anxious?
Can your child deal with unexpected situations?
Conduct Skills:
Does your child often loose control?
Is your child often angry?
Does your child often start to argue?
Does your child like to annoy others?
Hyperactivity:
Can your child sit still for longer?
Does your child often fidget?
Does your child often act without thinking?
Is your child concentrated when having started something?
Peer Relationships
Does your child like to meet new children?
How well does your child get along with friends?
How well does your child get along with classmates?
Grades:
How good is your child in Math
How good is your child in Reading
How good is your child in Writing

Note: These are the questions on children's development used in the GCP survey. Mothers were asked to report the performance in the school subjects math, reading, and writing. Additionally, they had to answer detailed questions on their child's behavior we grouped into the four categories of the strength and difficulty questionnaire (SDQ). Mothers could answer the questions based on a scale from 1 to 4 with 1 indicating "agree strongly" and 4 indicating "disagree strongly".

Table 2.10: Daily Activities

	Pooled	No Afterschool	Afterschool	Afterschool vs. No Afterschool	
	N=857	N=689	N=168	Difference	Std. Error
Panel A. Family Life					
Having lunch together with the family	0.167	0.185	0.096	-0.089***	(0.032)
Having dinner together with the family	0.541	0.536	0.565	0.030	(0.043)
Panel B. Hobbies					
Club (Sports, Dancing, etc.)	0.755	0.763	0.720	-0.043	(0.037)
Private Lessons (Music, Sports, etc.)	0.403	0.419	0.339	-0.079*	(0.042)

Note: The statistics are based on the first period of the GCP survey. All variables are binary variables. Standard errors in parentheses.
* p<0.10, ** p<0.05, *** p< 0.01.

Table 2.11: Detailed Skill Development

	Pooled	No Afterschool	Afterschool	Afterschool vs. No Afterschool
	N=857	N=689	N=168	Difference
School Performance				
Grades	0.093	0.127	-0.045	-0.172**
	(1.003)	(0.998)	(1.014)	(0.086)
Socio-Behavioral Development				
SDQ Score	-0.040	0.022	-0.292	-0.314***
	(1.004)	(0.992)	(1.016)	(0.086)
Components of the SDQ Score:				
Emotions	-0.049	-0.013	-0.196	-0.182**
	(1.021)	(0.999)	(1.100)	(0.088)
Conduct	-0.056	-0.017	-0.216	-0.199**
	(0.989)	(0.972)	(1.047)	(0.085)
Hyperactivity	-0.012	0.042	-0.233	-0.275***
	(1.021)	(1.017)	(1.010)	(0.087)
Peer Relations	0.010	0.033	-0.082	-0.115
	(1.021)	(0.999)	(1.104)	(0.088)

Note: The statistics are based on the first period of the GCP survey. Treated children are those children who attended after-school center-based care in the first period. Control children are those children who did not. Standard errors in parentheses.
* p<0.10, ** p<0.05, *** p< 0.01.

Table 2.12: Labor Market Participation of Parents

	Pooled	No Afterschool	Afterschool	Afterschool vs. No Afterschool	
				Difference	Std. Error
Mother is working	0.63	0.60	0.77	0.173***	(0.041)
	(N=857)	(N=689)	(N=168)		
Partner is working	0.94	0.95	0.90	-0.056***	(0.022)
	(N=780)	(N=644)	(N=136)		
Mother has full-time job	0.15	0.13	0.23	0.101**	(0.043)
	(N=388)	(N=304)	(N=84)		

Note: The statistics are based on the first period of the GCP survey. Treated children are those children who attended after-school center-based care in the first wave. Control children are those children who did not. Standard errors in parentheses.
* p<0.10, ** p<0.05, *** p< 0.01.

Table 2.13: Problems, Conflict Behavior and External Help

	Pooled	No Afterschool	Afterschool	Afterschool vs. No Afterschool	
				Difference	Std. Error
Panel A. Problems because of...					
	N=857	N=689	N=168		
housework	0.272	0.246	0.377	0.131***	(0.038)
(marital) partner	0.246	0.227	0.325	0.098***	(0.037)
stress in general	0.551	0.529	0.641	0.112***	(0.043)
parents' health	0.114	0.105	0.151	0.046*	(0.027)
behavior of the child	0.176	0.169	0.204	0.035	(0.033)
behavior of adults in family	0.140	0.140	0.138	-0.002	(0.030)
school of child	0.157	0.156	0.162	0.006	(0.031)
financial issues	0.162	0.156	0.186	0.029	(0.032)
alcohol	0.026	0.022	0.042	0.020	(0.014)
child's health	0.076	0.070	0.102	0.032	(0.023)
bad relations in family	0.108	0.101	0.138	0.037	(0.027)
household duties	0.269	0.254	0.329	0.075**	(0.038)
job-related stress	0.326	0.299	0.437	0.138***	(0.040)
lack of time	0.439	0.437	0.449	0.012	(0.043)
Panel B. Conflict Behavior					
Latest conflict with child was yesterday	0.443	0.438	0.471	0.033	(0.059)
	(N=485)	(N=400)	(N=85)		
Shouting at child in conflict	0.557	0.552	0.577	0.025	(0.043)
	(N=846)	(N=683)	(N=163)		
Ignoring child after conflict	0.082	0.085	0.067	-0.017	(0.024)
	(N=846)	(N=683)	(N=163)		
Panel C. External help					
	N=491	N=402	N=89		
Getting help from child psychologist or similar	0.283	0.269	0.348	0.080	(0.053)
Getting help from teacher in school or similar	0.240	0.236	0.258	0.022	(0.050)

Note: The statistics are based on the first period of the GCP survey. Treated children are those children who attended after-school center-based care in the first wave. Control children are those children who did not. Standard errors in parentheses.
* p<0.10, ** p<0.05, *** p< 0.01.

Table 2.14: Value Added Results: Detailed

	School Performance	*Socio-behavioral Development*
After-School Care	0.0638	0.0220
	(0.0757)	(0.0751)
Cognitive Skills	0.515***	0.0540*
	(0.0301)	(0.0298)
Non-Cognitive Skills	0.00773	0.500***
	(0.0297)	(0.0295)
Child is 7 years old	-0.00271	0.0657
	(0.0753)	(0.0748)
Child is 9 years old	0.00527	0.0196
	(0.0843)	(0.0837)
Child is 10 years old	0.145	-1.427*
	(0.819)	(0.814)
Child had low birthweight	-0.129	0.132
	(0.0893)	(0.0887)
Child has disorder	-0.00470	-0.0777
	(0.0744)	(0.0739)
Child is female	0.170***	0.237***
	(0.0579)	(0.0575)
Child's nationality is German	-0.0659	0.0722
	(0.108)	(0.107)
Child attended earlycare	0.0886	-0.0947
	(0.0776)	(0.0770)
Mother's education: primary	-0.186**	-0.0119
	(0.0760)	(0.0755)
Mother's education:university	0.0991	0.122
	(0.0810)	(0.0804)
Mother is single	-0.0938	-0.121
	(0.134)	(0.133)
Age of mother: under 30	0.161	-0.248*
	(0.129)	(0.128)
Age of mother: 30 to 34	0.0589	-0.186**
	(0.0836)	(0.0830)
Age of mother: 40 to 44	-0.0178	0.0363
	(0.0691)	(0.0687)
Age of mother: 45 to 49	0.00332	0.128
	(0.118)	(0.117)
Age of mother: 50plus	0.252	0.0162
	(0.339)	(0.337)
Child has siblings	-0.0666	-0.102
	(0.0871)	(0.0865)
Total net income 1250-2250 euro/month	0.113	-0.0450
	(0.142)	(0.141)
Total net income 2250-3250 euro/month	0.262*	0.0736
	(0.144)	(0.143)
Total net income higher than 3250 euro/month	0.314**	0.148
	(0.154)	(0.152)
Population Density	0.0000123	-0.0000112
	(0.0000570)	(0.0000566)
Unemployment Rate	-0.0474	-0.0362
	(0.0642)	(0.0637)
Female Employment Rate	0.0816	0.0423
	(0.0737)	(0.0732)
Net Migration	-0.00813	0.0153
	(0.0106)	(0.0106)
GDP p.c. in 1000euro	0.00281	-0.00479
	(0.00407)	(0.00405)
% of Children aged 6-10	0.106	-0.183
	(0.124)	(0.123)
Constant	-1.161	0.778
	(0.740)	(0.735)
State Dummies	YES	YES
State Dummies*Urban	YES	YES
N	857	857
r2	0.374	0.390

Notes: This table summarizes results of the value added regressions using the whole sample of 857 children. The first column shows the estimate for the regression of school performance in the second period on after-school care attendance in the first period, individual and regional characteristics in the first period, state dummies and state*urban interaction terms, and on both development measures in the first period. The second column shows the result of a corresponding regression with socio-behavioral development as dependent variable. Standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.
Source: Own calculations, based on the German Child Panel.

Table 2.15: Sensitivity Analysis: Using County Dummies

	School Performance	*Socio-behavioral Development*
Panel A: Pooled Sample: N=857	0.029	-0.021
Previous After-School Care Attendance	(0.079)	(0.078)
(168 treated, 689 control)		
Panel B: Education of Mother		
Primary Education: N=175	0.291	0.408*
Previous After-School Care Attendance	(0.253)	(0.230)
(27 treated, 148 control)		
Secondary Education: N=503	0.080	0.100
Previous After-School Care Attendance	(0.109)	(0.111)
(88 treated, 415 control)		
Uni Degree: N=156	-0.441**	-0.345
Previous After-School Care Attendance	(0.193)	(0.231)
(47 treated, 109 control)		
Panel C: Income of Family		
Low Income: N=284	0.127	0.220
Previous After-School Care Attendance	(0.142)	(0.133)
(66 treated, 218 control)		
Middle Income: N=297	0.133	-0.106
Previous After-School Care Attendance	(0.165)	(0.179)
(46 treated, 251 control)		
High Income: N=217	-0.165	-0.095
Previous After-School Care Attendance	(0.181)	(0.189)
(44 treated, 173 control)		

Notes: This table summarizes results of the value added regressions. Panel A shows estimates of the after-school care coefficient using the whole sample of 857 children. The first column shows the estimate for the regression of school performance in the second period on after-school care attendance in the first period, both development measures in the first period, individual characteristics in the first period, and a set of county dummies. Individual characteristics include characteristics of the child (age, gender, low birth weight, innate disorder, nationality, and early care attendance), the mother (age, education and single parenthood), and the family (number of siblings and net household income). The second column shows the result of a corresponding regression with socio-behavioral development as dependent variable. Panel B shows the respective regression results for children with respect to their mothers' education. Panel C stratifies with respect to the net monthly household income (categories that are used in the survey: below 2250 euro, between 2250-3250 euro, above 3250 euro).
* $p<0.10$, ** $p<0.05$, *** $p<0.01$
Source: Own calculations, based on the German Child Panel.

Chapter 3

From Dawn till Dusk – Implications of Full-Day Care for Children's Development

This chapter is joint work with Christina Felfe from the University of St. Gallen.

3.1 Introduction

Nowadays, more children than ever attend some kind of child care institution. Across all OECD countries, more than 80% of all three- to five-years-olds are enrolled in a child care program.[1] In most developed countries, this share even exceeds 90%. The intensity of child care, however, varies considerably across countries and rarely covers a full working day. Even in countries that exhibit full coverage – such as Germany, the Netherlands or the United Kingdom –, child care institutions mainly function on a half-day basis. The available child care institutions thus only partially facilitate the combination of market work and family life. To overcome this shortcoming, many

[1] All information on enrollment rates and opening hours are taken from the OECD Family Database (www.oecd.org/social/family/database, accessed on November 1, 2016).

OECD countries are currently debating whether to expand the opening hours of the existing child care institutions.[2] Yet, what are the consequences for children when attending a child care center on a full-day basis in contrast to attending it on a half-day basis? Answering this question is particularly relevant in light of the fact that attending child care on a full-day basis deprives children from spending valuable time with their primary caregiver – who in most cases is the mother.

This paper investigates to which extent prolonged opening hours of child care centers affect children's development. We study the effects of a substantial expansion of the number of full-day slots in one West German state over the last decade. This expansion was triggered by several reforms of the child care system, in particular in 1999 and 2005. These reforms occurred subsequently to a large expansion in child care centers in the late 1990s, which guaranteed a slot in child care on a half-day basis. Thus, our study analyzes the impact of moving from offering child care on a half-day basis to offering child care on a full-day basis. Importantly, the expansion of full-day slots did not occur at an equal pace, but the speed of the expansion varied considerably across municipalities. Our identification strategy exploits these municipality-specific deviations from the overall trend in the timing of the expansion. To be more precise, we rely on within-municipality variation in the supply of full-day slots over time net of the general trend in expanding the supply of full-day slots and net of any variation in the socio-economic and demographic composition of the municipalities over time. In addition, we consider possible adjustments in child care centers' structural quality when expanding the opening hours.

Our analysis uses two new and exceptionally rich data sources: First, we use administrative data from school entrance examinations in Schleswig-Holstein from 2004 to 2012. These data provide us with administrative records on children's development, in particular their overall school readiness, their motor skills and their socio-emotional maturity, for almost 100'000 children at school entrance. Second, we possess administrative information on child care centers which allow us to construct a series of

[2] For instance, Germany mandated an expansion of the opening hours of child care centers in the context of their Child Care Expansion Law in 2005, the Netherlands did so in the context of their Child Care Act in 2005, the United Kingdom in 2006 and British Columbia (Canada) in 2010.

structural quality indicators. The latter information may be particularly relevant to assess the role of care center quality when expanding care centers' opening hours.

Our empirical analysis leads to the following results: First, subsidized full-day care has negative effects on children's socio-emotional development: full-day separation from the primary caregiver – who is most likely the mother – entails problems for children's social maturity and emotional stability. On average, there are no effects on children's school readiness and motor skills. Second, subgroup analysis reveals that results are strongest for children from educationally disadvantaged backgrounds, immigrant ancestry and single parent households. Nevertheless, subgroup analysis also points towards some benefits in terms of school readiness: the more full-day slots are available, the more likely are children of immigrant ancestry to be assessed to be ready for school. Importantly, our results are robust to a series of alternative specifications.

Our study relates to a growing literature which investigates the effects of providing universal access to child care centers on children's skill acquisition and thus the effects of the extensive margin (providing child care versus not providing child care). Most of these studies base their identification strategy on a rapid expansion of the child care system. Findings from these studies are mixed and range from negative effects (Baker *et al.*, 2008), to neutral effects (Cascio, 2009; Datta-Gupta and Simonsen, 2010) and to positive effects (Magnuson *et al.*, 2007; Fitzpatrick, 2008; Gormley Jr. *et al.*, 2008; Berlinski *et al.*, 2009; Havnes and Mogstad, 2011; Noboa Hidalgo and Urzua, 2012; Felfe and Lalive, 2014; Felfe *et al.*, 2015; Cornelissen *et al.*, 2016). The reasons underlying these heterogeneous effects may range from methodological differences to contextual differences such as who are the compliers (which children react to the expansion), what are the counterfactual care modes or how the child care system is designed.

Extrapolating from the findings of this literature which focuses mostly on the effects of center-based care on a half-day basis is leading to wrong conclusions if the returns to time spent with alternative caregivers are non-linear. Studies investigating the intensive margin, and thus expansions of the opening hours of child care centers,

are scarce. There are some earlier studies providing correlations between full-day child care and child development (Cryan *et al.*, 1992; Gullo, 2000; Walson and West, 2004).[3] Only recently there are a few studies that aim at providing causal estimates for the impact of attending child care on a full-day basis. Rathburn and West (2004) and DeCicca (2007), for instance, rely on data from the Early Childhood Longitudinal Study and control for initial differences between children attending kindergarten (at age 5) on a half- and a full-day basis. None of these studies reveal any significant gains from attending child care on a full-time basis in terms of children's academic achievement. Cannon *et al.* (2006) exploit differences in state policies regulating the opening hours of child care centers and find short-lived positive effects on children's academic achievement. The study most closely related to ours is the one by Friesen *et al.* (2013). They exploit the staggered introduction of full-day kindergarten in British Columbia in Canada to analyze the impact of attending kindergarten on a full-day basis (at age 5). Their results reveal some short-lived negative impact of full-day kindergarten on children's development, in particular on parental reports of children's behavior and emotional development.[4]

Unlike previous studies, our main evaluation relies on administrative records instead of survey data. Administrative data have two main advantages over survey data: first, they allow us to mitigate the problem of misreporting – an issue which is likely to arise if parents justify their decision to enroll their child into child care on a full-day basis; second, they cover the full population of children belonging to a school cohort and thus circumvent the problem of attrition or non-response – problems likely to arise in the case of survey data (in particular, if the usage of full-day child care comes along with increased parental labor force participation).

The contribution of our study is threefold: First, we study the shift from a half-day to a full-day schedule in child care available to children age 0-6 years. Thus, in

[3] In addition, there exist several studies analyzing the impact of full-time maternal employment on child development (Waldfogel *et al.*, 2002; Brooks-Gunn *et al.*, 2002; Hill *et al.*, 2005). Similar to most existing studies on full-day kindergarten, these studies address endogeneity of maternal employment by controlling for a wide range of observable characteristics.

[4] There is one further study which contrasts the development of primary-school-aged children who attend after-school care with children who are taken care of by the mother in the afternoon hours (Felfe and Zierow, 2014), see Chapter 2 of this thesis. While this study does not find any effects on average, it finds beneficial effects for children from disadvantaged backgrounds.

contrast to previous studies, we study a much longer treatment (up to six years in contrast to only one year). As a consequence, resulting effects are potentially much stronger. Second, we simultaneously analyze a broad range of child development outcomes, including aspects of children's overall school readiness, motor skills and socio-emotional maturity. Finally, we draw on administrative data on child care centers' structural quality which allow us to assess the role of care centers' quality when increasing opening hours.

The reminder of this study is structured as follows: the next section describes the child care system in Germany as well as the reform under study. Sections 3.3 and 3.4 introduce the data and identification strategy, respectively. Section 3.5 presents our main findings as well as results from sensitivity and subgroup analysis. Section 3.6 reflects on potential mechanisms explaining the results. Section 3.7 finally concludes.

3.2 Background

3.2.1 The Child Care System in Germany[5]

Germany offers child care at two levels. Early child care is available for children between 0 and 3 years, while later child care – the so-called kindergarten – is available for children between 3 and 6 years. Since 1996 every child turning 3 years old is legally entitled to a slot in a child care center. As a result of this policy, since the early 2000s more than 90% of children entering school have attended kindergarten for at least 2 years.[6] In contrast, early child care is a rather recent institution in West Germany.[7] A legal claim on a slot in early child care exists since 2013, but even then only 23% of all

[5] This section draws on our own calculations of the statistics provided by the German Child and Youth Services ("Kinder- und Jugendhilfe") as well as on the official publications based on these statistics (Riedel *et al.* (2005), Lange *et al.* (2007), Huesken (2010) and Strunz (2011)).

[6] In West Germany in 2006, for instance, 75% among 3-year-old children attended a child care center, while 92% of all 4-5-year-old children did so. In 2011, the respective shares amounted to 85% and 95%.

[7] As a heritage from the former German Democratic Republic, in East Germany universal child care is available for all ages, also at a full-day basis.

West German 0-to-3-year-old children made use of it. In prior years, attendance rates were much lower: up to 2002 less than 3% of all 0-to-3-year-old West German children attended a care center, in 2006 attendance rates rose to 7% and in 2011 attendance rates amounted to 16%. Thus, child care centers are mainly an institution for 3-to-6-year-old children.

Child care is mostly organized in child care centers.[8] Care centers are run by subsidized non-profit organizations, such as the municipality, the church or welfare organizations.[9] Subsidies come from three public entities: the state usually pays a large amount of the total operating costs. Schleswig-Holstein, the state under study, pays an annual amount of 60-70 million Euros which corresponds to approximately 15% of total operating costs. This money is distributed to the counties according to the number of children enrolled in care centers, the number of immigrant children enrolled in care centers and the opening hours of child care centers. Counties augment this amount by further 5% of the operating costs. The largest share of the operating costs is borne by the municipalities (around 40% of operating costs). The remaining 40% of the operating costs are paid by private organizations (10%) and parents (30%). Parental fees are differentiated according to family size, the number of siblings enrolled in child care and family income. The costs for a half-day slot range between 0 and 200 Euro/month and for a full-day slot between 0 and 420 Euro/month plus a separate fee for lunch (around 80 Euro/month). As a consequence of the large subsidization of child care centers, privately arranged non-parental care is very uncommon in Germany. Families primarily use public center-based care during their work hours.[10]

States are in charge of regulating the quality of center-based care in Germany.[11] Regulations concern aspects such as opening hours, group size, staff-child ratio, but

[8] Since the child care expansion law in 2005, extra-familial childminders have gained increasing importance. However, in West Germany in 2006 only 1.2% of all 0-2-year-old children have been taken care of by a childminder, in 2011 only 3.7%. Among 3-6-year-old children this share is negligible (in 2011: 0.5%).

[9] Only a negligible share is run by a private provider (in West Germany in 2006 the private share was at 1%).

[10] Less than 5% of all families with children below the age of 6 years use private child care.

[11] Notice that every institution – independent of being run by the church, a welfare organization or the municipality – has to adhere to these regulations

also staff qualifications. On average, care centers have to remain open for at least four hours on five days per week. Regulations regarding groupsize and staff-child ratio in the case of 3-6-year-old children are as follows: Playgroups can have at most 25 children and need to be supervised by at least one certified child care worker and one or two assistants. The degree of a child care worker requires two years of theoretical training and at least two years of practice in a child care center. Care centers comply with these regulations: in 2006 playgroups accommodated on average 20 children between 3 and 6 years. 62.3% of the employed staff had a child care worker degree. Overall, about ten children were supervised by one staff member. Moreover, in line with the minimum required opening hours, in West Germany in 2006 53% of children attending a care center were taken care off on a half-day basis, while 47% had a full-day slot.

Subsequently to the mandate in 1996, which entitled every child between 3 and 6 years to a slot in a care center, many German states revised their child care laws. Schleswig-Holstein, the state under study, did so in 1999 (in its law on child care centers, the so-called Kindertagesstättengesetz, or short KiTaG). Revisions concerned mainly expansions of existing child care centers to accommodate also younger children (between 0 and 3 years old) and to offer longer opening hours. In 2005, the German government enacted the day care expansion law to deal with the remaining shortages in the German child care system (Tagesbetreuungsausbaugesetz, §24(1), SGB VIII). Besides kicking off the expansion of slots available to 0-to-3-year-old children, this mandate triggered a strong expansion of full-day slots. Figure 3.1, Panel A illustrates this expansion for the region under study. In 1998, only 30% of all children enrolled in a care center were offered a full-day slot.[12] The revision of the KiTaG led to a slight increase in this share. Following the 2005 mandate this share increased remarkably: in 2006 more than 40% of all children enrolled in preschool were offered a full-day slot, in 2008 around 45% and in 2011 more than 50%. Figure 3.1, Panel B shows additionally the expansion of early care. The peak of the expansion in early care occurred slightly later, in particular from 2008 onwards. The underlying reason is the additional pressure to expand slots in early care due to the law on support for

[12] Register data on child care centers are only available from 1998 onwards. We therefore cannot provide any previous trends on the supply of slots in child care centers.

children in 2008 which announced the legal claim on a slot in early care from 2013 onwards.

Who is responsible for expansions in the child care system? In Schleswig-Holstein, the local organizations of the Child and Youth Services are obliged to annually assess the demand for child care slots and desired opening hours (§7, KiTaG). Key figures for the prediction of the demand for longer opening hours are the number of preschool-aged children and the number of mothers working or desiring to work. Based on the predicted demand and the existing supply, the Child and Youth Services then predict the required expansion and set a realistic time horizon for implementation. Limiting criteria for expanding the opening hours of child care centers are mainly a lack of appropriate infrastructure and qualified staff.

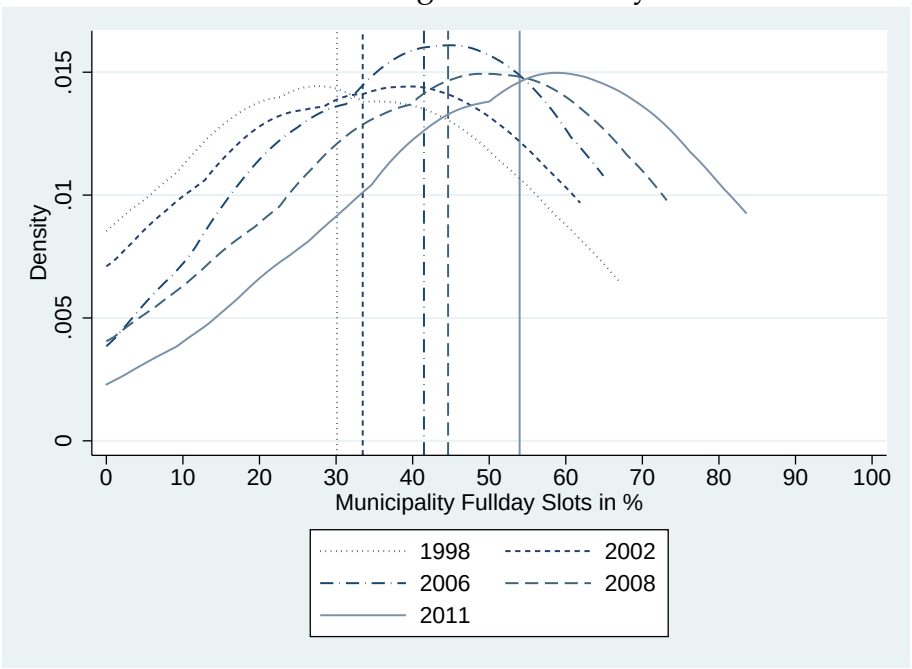
Who bears the additional costs? As described above there is some financial aid coming from the state as well as the federal government. Besides the direct transfer from the state of Schleswig-Holstein to child care centers, there are indirect transfers via equalization transfers between the municipalities within each state (regulated in the so-called *Finanzausgleichsgesetz*): municipalities are allowed to go into debt to venture bigger investments related to their child care system; indebted municipalities are then indirectly reimbursed for these investments through equalization transfers running from richer to poorer municipalities. Since 2009, the federal government additionally supports the creation of full-day slots via the financial stability act which dedicates money directly to the maintenance and expansions of existing child care centers.

3.2.2 School Entrance Examination

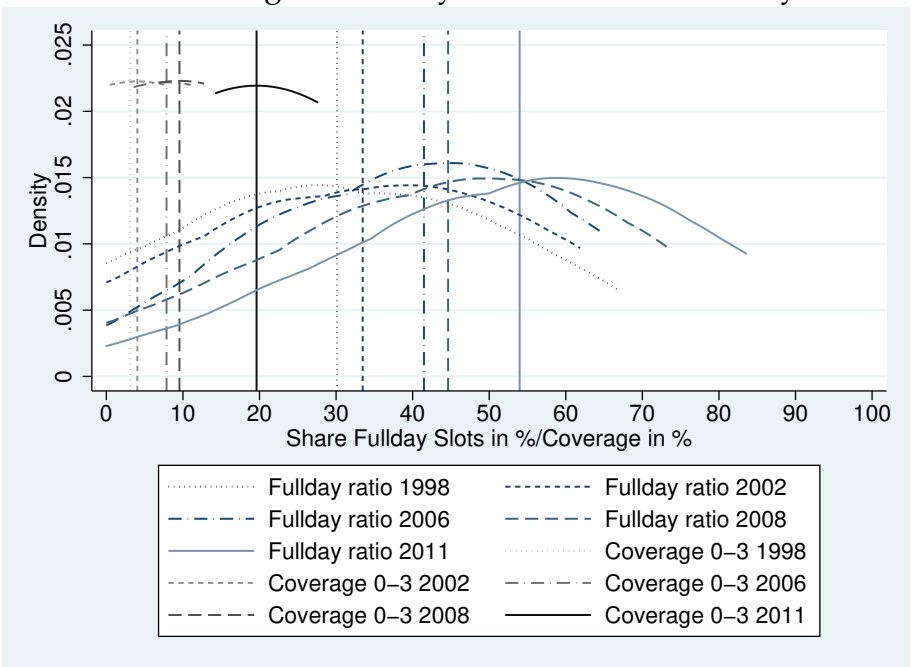
In Germany, children undergo several mandatory medical screenings between birth and primary school. These medical screenings promote children's health by diagnosing medical anomalies and providing necessary treatment as early as possible. An important medical screening is the school entrance examination (SEE). Besides documenting a child's health, the focus of this examination is to determine whether a child is "ready" to follow the primary school curriculum or not.

Figure 3.1: Expansion of the Child Care System

Panel A: Coverage rate: Full-day slots



Panel B: Coverage rate: Early Care Slots and Full-day slots



Notes: Panel A plots the share of full-day slots over all slots in care centers available to 0-to-6-year-old-children for the years 1998, 2002, 2006, 2008 and 2011 in the examined area in Schleswig-Holstein. Panel B plots additionally the coverage rate with slots in center-based care among 0-to-3-year-olds.

Source: Statistics of the Child and Youth Services in Germany 1998-2011. Own calculations.

In the context of the SEE, pediatricians employed by the local health service examine all children in the year prior to entering primary school and thus when children are around 6 years old. They provide a medical diagnosis for several dimensions of children's state of development, among others, children's motor skills and socio-emotional maturity. The motor skills diagnosis concerns coordination and motor capacities of the child. Children have to stand on one leg, jump on one leg, jump left and right for a longer time span. The socio-emotional maturity assessment is based on the pediatrician's observations as well as on a questionnaire designed to identify emotional problems, behavioral problems, hyperactivity, peer relationships, and social behavior given to the accompanying caregiver: the well-established *Strengths and Difficulties Questionnaire*, short SDQ (Goodman, 1997).

One might worry that the assessment of socio-emotional maturity might be affected by subjective perceptions of the caregiver, or by non-response problems. Considering that the pediatricians re-assess children's socio-emotional maturity and that in 93% of all cases a medical diagnoses regarding socio-emotional maturity is available, reporting bias and non-response bias are not a major concern in our context. One further concern may be that pediatricians are subjective in their assessment (and possibly consider children's family and institutional environment in their assessment). This, however, is unlikely to occur and can be accounted for by the municipality fixed-effects, which given the rather low turnover of pediatricians implicitly corresponds to a pediatrician fixed-effect (on average every pediatrician is assigned to 5-8 municipalities).

The school readiness assessment evaluates whether a child is ready for school or not. This assessment is correlated with the medical diagnoses, but the correlations need not to be perfect. Pediatricians weigh the diagnoses concerning the different developmental domains and may include further aspects, i.e. proficiency in German or cultural assimilation for immigrant children. The overall diagnosis thus needs not to be the sum of the different skill dimensions. Even if the school readiness diagnosis is not binding, it is an important piece of information regarding school enrollment and can be crucial for parents' decision whether to enroll their child in school or not.

Why should economists care about children's performance in the dimensions assessed in the school entrance examinations? There is a growing amount of research relating such early performance indicators to later success on the labor market. Gregg and Machin (1999, 2001), for instance, discuss the relevance of children's early cognitive abilities for their later success in the labor market. Duncan *et al.* (2007) show that dimensions assessed in the school entrance examination – such as intellectual skills and socio-emotional maturity – are key in predicting later educational achievements. Finally, motor skills are surprisingly very strong predictors for children's later achievements (Grissmer *et al.*, 2010).

3.3 Data and Descriptive Statistics

3.3.1 Data

Administrative records from *Schleswig-Holstein*'s school entrance examinations (SEE) are our main data source. The data is structured in school entrance cohorts. A school entrance cohort comprises all children who turn 6 years old between July of the previous year and June of the same year as school entrance.[13] This study draws on data for the school entrance cohorts 2004 to 2012 and thus on birth cohorts 1997-2006. As described in the previous section, SEE data contain medical assessments on, among other dimensions, children's motor skills and socio-emotional maturity. The medical diagnosis can take five forms: "normal development", "some problems, but no treatment is necessary", "some problems, already in treatment", "problems, treatment necessary", and "problems which will reduce the child's performance in school". Based on this diagnosis, we construct a binary indicator for each of the two dimensions assessed in the SEE (motor skills and socio-emotional maturity), which is equal to one if the child does not exhibit any problem in the assessed tasks. The SEE

[13] Children who were not ready for school in one year undertake a special examination one year later and thus are not included in the baseline SEE. Parents whose children turn 6 years between July and December of the same year can ask their child to be examined a year before the official SEE would have taken place. We exclude these children from our analyses.

provides us furthermore with the pediatrician's assessment whether a child is ready to follow the school curriculum or not. The recommendation can take the following three forms: "ready for school", "school enrollment only with support provided by the teacher", and "special needs education required". We construct again a binary indicator which equals one if the child is ready for school.

The SEE also contains parental reports on child and family background. Among other questions, parents indicate whether their child attended child care. Yet, they do not provide any information on the amount of hours their child attended child care. In other words, we do not possess any direct information on the intensive margin on an individual basis – i.e. whether the child attended care on a full-day basis. Instead we rely on the average rate of full-day slots among all slots available in care centers on the municipality level – provided by our second data source described in turn. Thus, similar to previous studies investigating the impact of universal child care on children's development (Baker *et al.*, 2008; Fitzpatrick, 2008; Cascio, 2009; Havnes and Mogstad, 2011; Felfe *et al.*, 2015), we can only provide estimates for the intention-to-treat effect (ITT), but not for the treatment effect itself.

Administrative records of all child care centers are our second data source (the so-called *Kinder- und Jugendhilfestatistik* or *Statistics on Child and Youth Services*). These records contain detailed information on the provider, the number of children enrolled and the staff employed and thus, allow us to construct the following series of indicators describing the care centers: coverage rates among 0-to-3-year-old children and 3-to-6-year-old children, the share of full-day and half-day slots, the provider (public provider vs. the church vs. other providers, which are mostly welfare organizations), as well as structural quality features such as group size and the staff composition in terms of age, gender, pedagogical degree and workload. All information is available at the care center level. Since a substantial share of care centers host children of different age groups[14], we cannot distinguish between slots offered to 0-to-3-year-old children and slots offered to 3-to-6-year-old children. As such our treatment – the share of full-day slots – as well as any other information on care centers refer to children at the age of 0 to 6 years. Nevertheless, as pointed out before, until 2006 only up to

[14] In 1998 25% of care centers hosted children of different age groups, in 2011 55%.

6% of all slots in care centers are offered to 0-to-3-year-old children, while more than 93% of all slots in care centers are offered to 3-to-6-year-old children. In other words, the focus of our analysis lies on the effect of expanding full-day care which is mostly available to 3-to-6-year-old children.

The fact that information on our treatment – the share of full-day slots – is only available at the care center level (being theoretically accessible for 0-to-6-year-olds), rises the question of when to measure the share of full-day slots. In other words, we have the choice of measuring the supply of full-day slots at any age between zero and six. To circumvent any endogeneity of care center features to the parental decision of enrolling their child in child care and in particular, in child care on a full-day basis, we choose as treatment variable the share of full-day slots available in children's birth year and thus prior to children's own enrollment in child care. Yet, we test the robustness of our results when choosing alternative points in time to merge the share of full-day slots with child outcomes (see Section 3.5.2).

The smallest regional level available in both data sources is the municipality.[15] Data protection issues, however, restrict the number of municipalities available for scientific research.[16] We possess identifiers for 75 municipalities (belonging to 8 out of 15 counties) which allow us to merge the available administrative data on child care centers to the SEE data. We additionally merge information on the demographic and socio-economic composition of the municipalities via the municipality identifiers. These additional data are mainly part of the *"INKAR-Raumordnungsdaten"*, a data set on municipality characteristics published by the *Federal Institute for Research on Building, Urban Affairs and Spatial Development*; data on female full-time employment are provided by *"Statistikamt Nord"*, the statistical office of the North German states

[15] We observe the municipality of residence of a child at the SEE date. As post-birth mobility is low in West Germany, the municipality of residence at the SEE date is likely to be the same as the municipality of residence when children attend center-based care for most children in our sample.

[16] First, administrative data on care centers are only released if municipalities contain at least three care centers, otherwise only averages of care centers in neighboring municipalities are available. Second, not all municipalities can be identified in the SEE data. In fact, counties – the second smallest regional level in Germany – are in charge of gathering the results of the SEE and delivering them to the respective state office (which is the *Ministry of Social Affairs, Health, Family and Equality* in Schleswig-Holstein). When delivering the data to the ministry, counties have the right to anonymize municipalities and some of them do so.

Hamburg and Schleswig-Holstein. Restricting our sample to children for whom we possess information on all assessed dimensions – school readiness, motor skills and socio-emotional maturity – leads to a sample of 93,570 children belonging to nine school entrance cohorts and residing in 75 municipalities.

3.3.2 Descriptive Statistics

Table 3.1 provides descriptive statistics for the estimation sample which contains 93,570 children entering school between 2004 and 2012 (and thus born between July 1997 and June 2006). All child development dimensions are constructed as binary variables (see Section 3.3.1 for details).[17]

How do children perform in the various dimensions assessed in the SEE? As we can see in Panel A, 82.5% of all children are assessed to be socio-emotionally mature. 81.2% of all children exhibit motor skills which are age appropriate. Finally, 86.9% of all children are assessed to be ready for school, i.e. able to follow the curriculum taught in primary school. While children perform on average quite high in all dimensions assessed in the context of the SEE, the rather high standard deviation (ranging between 0.34 and 0.39) points towards strong heterogeneity in the performance of children at the time of school entrance.

Panel B reports some information on the characteristics of children and their family background. On average children are 73.8 months (6.1 years) old, around half of them are boys (52.4%), 12.5% are immigrants, 15.1% live with one parent only, and on average they have one sibling. Around one fifth of all children live in a family where the mother has a primary school degree (19.9%), one third of all children grow up in a family where the mother holds an intermediate education degree (32.2%) and more than a quarter where the holds a higher education degree (27.4%). The educational background for the remaining mothers is missing.

[17] Binary variables are marked with a (D) in Table 3.1.

Table 3.1: Descriptive Statistics

	Mean	SD
Panel A:Development Dimensions		
Socio-Emotional Maturity (D)	.825	.38
Motor Skills (D)	.812	.391
School Readiness (D)	.869	.338
Panel B: Child characteristics		
Age (in month)	73.761	3.888
Male (D)	.524	.499
Immigrants (D)	.125	.331
Birth weight (in gram)	3275.256	815.235
Birth weight missing (D)	.029	.167
Single parent (D)	.151	.358
Single parent: missing (D)	.063	.243
Nr of siblings (excl. kid)	1.023	1.047
Siblings: missing (D)	.136	.343
Mom's education: basic (D)	.199	.4
Mom's education: intermediate (D)	.322	.467
Mom's education: high (D)	.274	.446
Mom's education: missing (D)	.173	.378
Panel C: Care center characteristics		
Fullday Share	.336	.198
Coverage 0-3 years old (in %)	4.538	2.543
Coverage 3-6 years old (in %)	82.139	6.683
Share of public providers (in %)	20.944	16.158
Share of other providers (in %)	37.387	17.974
Share of church providers (in %)	41.669	18.664
Children per group	19.969	3.284
Age of staff (years)	38.179	2.328
Share of male staff (in %)	4.355	3.281
Share of staff with pedagogical degree (in %)	62.113	9.51
Share of full-time staff (in %)	35.813	14.678
Panel D: Regional Characteristics		
Citizens per km2	1021.08	700.396
Share 0-6-y-old children (in %)	5.796	.804
Votes for CDU and FDP in %	42.035	4.28
Votes for SPD in %	44.436	3.73
Votes for other parties in %	13.529	3.034
Log of GDPpc (in 1000Euro/Citizen)	3.211	.226
Local business tax rate	3.608	.49
Local tax rate on agrarian real property	3.172	.47
Local tax rate on other real property	3.573	.823
Employed female (in %)	44.251	2.66
Share fulltime-employed female (in %)	63.978	5.343
N=93570		

Notes: Individual descriptives are based on the full sample (2004-2012) of the School entrance examination data. All macro variables (at the center level and at the regional level) are measured in the year when children are born (1998-2006). Binary variables are marked with a (D).

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

Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord. Own calculations.

How does the provision with and the quality of child care in the region under study look like? Panel C displays the features of the child care centers located in the municipalities contained in our estimation sample, measured as the respective average at the municipality level of the year the child is born (and thus across the years 1998 until 2006). As described in Section 3.2, while the supply of slots in kindergarten is basically fulfilling the legal mandate of full coverage (82% of all children age 3-6 years old can attend kindergarten), early child care coverage is very low (4.5% of all children can attend early care). Yet, as shown in Figure 3.1 the coverage of early care has steadily grown over the last two decades and reaches 23% in 2013. Regarding the share of available full-day slots, we observe the following: on average, one third of all children is offered a full-day slot (34%). Again this share is steadily rising over the last two decades (see Figure 3.1) – a fact we base our identification strategy on (see Section 3.4 for details): in 1998, the share of full-time slots amounts to 30.1%, in 2006 the share of full-time slots is already at 41.5% and in 2011 it equals 53.9%.

Most child care centers are either run by the church (41.7%) or by a welfare organization (37.4%), but still a fifth of all care centers are run by the municipality (20.9%). On average there are 20 children in one group, the staff working in the child care centers is on average 37.4 years old, most are female (95.4%), the majority holds a pedagogical degree (62.1%), and around a third works full-time (35.8%).

Panel D finally provides some information on the municipalities the children live in (again merged to the SEE data for the year when children are born and thus measuring the average across the years 1998-2006). Female employment, the share of women working full-time, GDP per capita, the local tax rates[18] and the percentage of children between 0 and 6 years are comparable to the West German average. The region under study is a bit more densely populated – notice that four of the counties under study are urban counties – and less conservative – the vote share for the conservative parties is substantially lower than the West German average.

[18] In Germany, municipalities have the legal right to annually decide on the tax rate of three different kinds of local taxes: business taxes, taxes on agrarian real property and taxes on other real property.

3.4 Identification

To estimate the causal effect of the supply of full-day slots in child care centers on children's development we rely on within-municipality variation in the supply of full-day slots. Our empirical specification looks as follows:

$$Y^s_{imc} = \beta F_{mc} + \gamma C_{mc} + \delta Z_{mc} + \eta X_i + \mu_c + \psi_m + \epsilon_{imc} \tag{3.1}$$

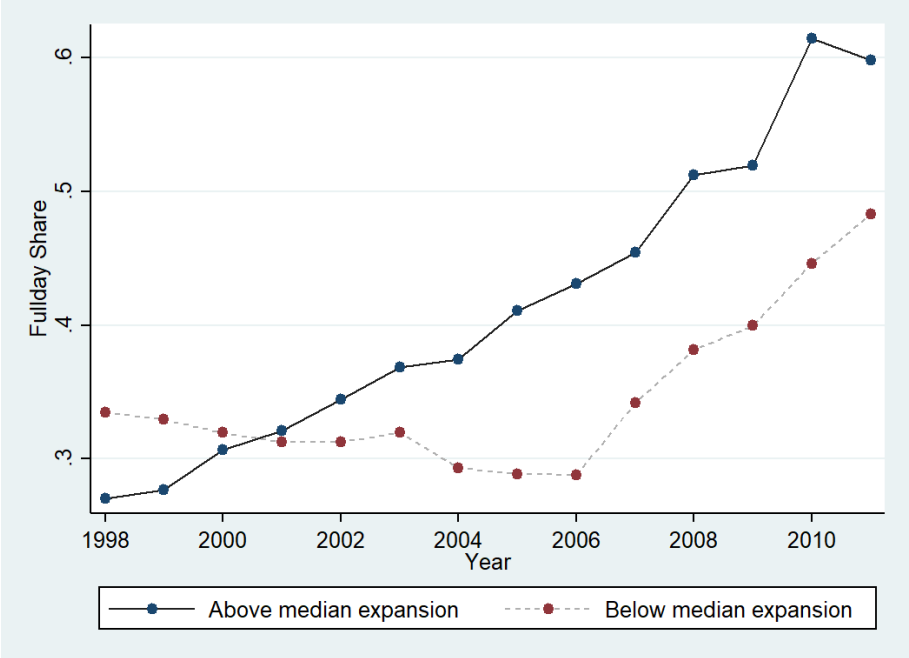
where Y^s_{imc} denotes skill dimension s of child i residing in municipality m and belonging to cohort c, F_{mc} stands for the share of full-day slots available to children belonging to cohort c and residing in municipality m, C_{mc} represents a set of care center features, Z_{mc} a set of municipality features, and X_i a set of individual background characteristics. The set of municipality dummies ψ_m allows us to control for the average level of full-day slots in the municipality, but also for any time-constant features of municipalities that may correlate with the timing of the expansion in full-day slots and with changes in children's development. By controlling for a set of cohort dummies μ_c, we abstract furthermore from the overall trend to expand full-day slots and exploit the variation in the timing of the expansion across municipalities. Finally, ϵ_{imc} represents an idiosyncratic shock.

The key identifying assumption is that F_{mc} is conditionally independent of the unobserved determinants of children's development ϵ_{imc}. Ideally we would rely on exogenous supply shocks of full-day slots. Instead we rely on an expansion of the supply of full-day slots triggered by revisions of the child care law and exploit variation in the timing of expansions across municipality. In other words, our identification strategy relies on within-municipality variation of the supply of full-day slots. Two questions may arise when relying on this identification strategy: first, is there sufficient variation in the supply of full-day slots? Second, is the timing of expansions in the supply of full-day slots indeed exogenous within municipality?

A first hint for the existence of sufficient variation gives Figure 3.1 in Section 3.2.1. It shows the distribution of the full-day share over time and helps getting a picture of the characteristics necessary for our analysis: there is variation of the share of full-day slots over the years (the mean is growing from 30% in 1998 to 41% in 2006) and at the same time there is high variation in every year across regions.

Further evidence for the expansion of full-day slots varying by starting time and intensity across the 75 municipalities provides Figure 3.2. We divide the municipalities in two groups: on the one hand, those expanding their full-day slot supply by more than median expansion and on the other hand, those expanding their full-day slot supply by less than median expansion. As Figure 3.2 shows, during the period under study (1998-2006) we observe differential time trends – there are quick and slow expanders –, but in the long-run they all aim at an expansion (the figure shows a lense shape relationship). Turning to the second major threat to our identification strategy, in the following we provide some information on the timing of the expansions in the supply of full-day slots. Expansions in full-day slots are determined by the demand of citizens for full-day slots and the ability and willingness of the providers and ultimately the municipality – as the municipality is stemming a large part of the costs – to expand the opening hours of child care centers. Proxies for the demand are the number of children eligible for child care (thus children who are 0-6 years old) and the share of employed women as well as the share of full-time employed women. Proxies for the ability of a municipality to expand full-day care are the economic standing of the municipality, which we proxy by GDP per capita, and the population density. One might be concerned that municipalities may cut down on other social expenditures or levy more taxes to finance the expansion of full-day slots. Individual taxes or social benefits, however, are set on the federal level. Thus, it is unlikely that the financial burden due to an expansion of the child care system is passed on to the citizens or crowds out other social expenditures. Municipalities, however, levy corporate taxes as well as taxes on real property and thus the financial burden may be passed on to the local economy. Yet, it is unlikely that municipalities are willing to damage its attractiveness as a business location in order to satisfy the demand for more full-day care. Municipalities may rather accept to go into debt and

Figure 3.2: Expansion of the Share of Full-Day Slots of Below- and Above-Median Expanders



Notes: The share of full-day slots is expressed as municipality mean. The group of above-median expanders includes those municipalities that experienced an above median growth between 1998-2006. The group of below-median expanders includes the other municipalities.
Source: Statistics of the *Child and Youth Services in Germany*, Own calculations.

make use of equalization transfers provided to poorer municipalities. In addition, as described in Section 3.2, state subsidies are relatively larger for child care centers running on a full-day basis. Nevertheless, we include the local business tax rate and the tax rates on both agrarian and other real property. Those rates are determined by the municipalities annually and provide an instrument to increase their financial capacities. Finally, to proxy for the willingness of a municipality to expand the supply of full-day slots, we additionally control for the election results in the last federal elections – there are clear differences in the party programs regarding the expansion of child care centers between the more conservative parties (CDU and FDP) and the less conservative parties (SPD and Grüne).

Besides financial consequences of the expansions of full-day care, one may worry about concessions in care centers' quality for the following reasons: i) one way to finance more/longer child care is to sacrifice quality; ii) the more rapid the expansion, the higher the probability that quality may suffer. For instance, it may be difficult to recruit adequately trained staff or groups may be joined in the afternoon hours. We account for potential adjustments in care centers' quality by controlling additionally for a set of structural quality parameters.

To shed some light on the driving forces behind the expansions in full-day care, we run a supplementary regression where the dependent variable is each municipality's full-day share expansion between 1998 and 2006 and the independent variables are the set of municipality and care center features in 1998. Results in Table 3.2 show that especially those municipalities that had a low share of full-day slots in 1998 increased their supply in the following years (i.e. those that had to "catch up"). Additionally, regions with a higher share of non-clerical institutions and a higher vote share for the Green party in 1998 experienced a larger expansion (i.e. those with less conservative attitudes regarding child care). Moreover, also municipalities with a higher share of full-time staff in 1998 expanded their full-day slot supply more than others in the following years (i.e. those who had fewer constraints on the supply side of care).

Table 3.3 provides yet another perspective and shows which regional characteristics change simultaneously with the share of full-day slots over time.[19] In line with the revisions of the child care law we observe a strong trend in the share of full-day slots over time: in particular, the school entrance cohort 2012 (i.e. birth cohort 2005/06) faces a 23 ppt higher supply of full-day slots than the school entrance cohort 2004 (i.e. birth cohort 1997/98). Importantly, there is no sign that municipalities simultaneously plan the expansion of full-day sots and the expansion of slots in early care. There is also no indication that expansions in full-day care are correlated with changes in care provided to 3-6-year-olds. The p-value resulting from an F-test for overall significance of coverage in early care and kindergarten is 0.765. The expansion in the share of full-day slots is also uncorrelated with basically all observable municipality features. The only exception is the political orientation of the municipality which can reflect both the attitude of the citizens as well as the priority given to child care by the political party. To be precise: in regions, where the citizens are more likely (by 1 ppt) to vote for a conservative party (either the CDU or the FDP), the share of full-day slots is likely to increase by 2 ppt less, in contrast to when the citizens are more likely to vote for the socialist party. Finally, there is no clear evidence for concessions in care centers' quality when expanding the opening hours. One exception is – not surprisingly – that care centers running on a full-day basis require more staff working on a full-time basis.

Given this evidence, the only remaining threat to identification are municipality-specific changes over time that are orthogonal to the municipality and care center characteristics we already control for and that affect children belonging to one cohort and living within a given municipality differentially. Yet, it is hard to come up with any systematic difference within a municipality over time that is likely to drive the expansion of full-day slots and affects children's development beyond the conditions we already account for. Finally, we are unaware of any further reform of the education system that might differentially affect children belonging to different cohorts. Yet, if there were one, the educational system is regulated at the state level and thus any

[19] In an alternative regression, shown in Table 3.9 we control for all municipality features in the previous year. This accounts for possible planning of the expansion of full-day slots. Results are basically unchanged.

change should be controlled for by the set of cohort fixed effects. Therefore, we are confident that the observed variation in timing of the expansion of the share of full-day slots is – conditional on the controls presented above – due to unplanned, non-systematic delays, e.g. delays in the construction of full-day care facilities, in the search for employees, in administrative decision-taking.

To sum up, we use within-municipality variation in the supply of full-day slots to identify the effect of longer child care center opening hours on child outcomes, in other words we look at differences in child outcomes due to different timing in full-day slot creation. We assume that the timing is exogenous when controlling for regional and care center characteristics that proxy the ability and willingness of municipalities to expand their full-day share.

3.5 Results

3.5.1 Baseline results

What is the impact of extending the opening hours of child care centers from half-day basis to full-day basis on children's development? Table 3.4 shows the results of estimating the empirical model (3.1) by OLS and clustering the standard errors at the municipality level.

The observed expansion of full-day slots in the region under study – which amounts to 11 ppt or 11 out of a 100 slots between the school cohort 2004 (born 7/1997-6/1998) and the school cohort 2012 (born between 7/2005 and 6/2006) – does not affect children's overall school readiness or motor skills on average. Yet, the observed expansion of full-day slots leads to a deterioration of children's socio-emotional maturity by 2.44 ppt, or 0.064 standard deviations (henceforth sd). Table 3.5 shows the results of our baseline specification when stratifying the sample according to the following characteristics: gender (boys versus girls), ethnic background (native versus immigrant), parental education (both parents possess of a university-entrance

Table 3.2: Determinants of the Expansion of Full-Day Slots

	Expansion of the Fullday Share 1998-2006
Fullday Share in 1998	-0.979***
	(0.103)
Coverage 0-3 years old (in %) in 1998	-0.018
	(0.020)
Coverage 3-6 years old (in %) in 1998	0.006
	(0.006)
Public provider(in %) in 1998	0.001*
	(0.001)
Other provider (in %) in 1998	0.001*
	(0.001)
Staff: fulltime (in %) in 1998	0.005***
	(0.001)
Children per group in 1998	-0.004
	(0.003)
Age of staff (years) in 1998	-0.007
	(0.006)
Staff: male (in %) in 1998	0.002
	(0.002)
Staff: pedagogical degree (in %) in 1998	-0.000
	(0.001)
Citizens per km2 in 1998	-0.000
	(0.000)
Fulltime-Employed female (in % of total working) in 1998	0.001
	(0.001)
Employed female (in %) in 1998	0.002
	(0.004)
Share 0-6-y-old children (in %) in 1998	-0.011
	(0.016)
Votes for CDU and FDP (in %) in 1998	0.022
	(0.018)
Votes for other parties (in %) in 1998	0.044
	(0.029)
Log of GDPpc (in 1000Euro/Citizen) in 1998	-0.002
	(0.010)
Local business tax rate in 1998	0.205
	(0.130)
Local tax rate on agrarian real property in 1998	-0.092
	(0.112)
Local tax rate on other real property in 1998	0.003
	(0.087)
County FE	No
Municipality FE	No
Joint significance - Slots (p-Value)	0.106
Joint significance - Institution (p-Value)	0.061
Joint significance - Center (p-Value)	0.000
Joint significance - Regional (p-Value)	0.022
Adj. R2	0.779
Children	93570

Notes: This table shows the coefficients of the OLS estimates of the expansion of full-day slots between 1998-2006 on regional characteristics in 1998. Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.
Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord. Own calculations.

Table 3.3: Which regional characteristics change simultaneously with the full-day share?

	Full-Day Share at Birth
Cohort 2005 (D)	0.015**
	(0.007)
Cohort 2006 (D)	0.031**
	(0.015)
Cohort 2007 (D)	0.049**
	(0.021)
Cohort 2008 (D)	0.098***
	(0.029)
Cohort 2009 (D)	0.119***
	(0.030)
Cohort 2010 (D)	0.139***
	(0.031)
Cohort 2011 (D)	0.215***
	(0.054)
Cohort 2012 (D)	0.229***
	(0.057)
Coverage 0-3 years old (in %)	0.001
	(0.008)
Coverage 3-6 years old (in %)	-0.002
	(0.003)
Public provider(in %)	-0.000
	(0.001)
Other provider (in %)	-0.001
	(0.001)
Children per group	-0.003
	(0.004)
Age of staff (years)	-0.008
	(0.007)
Staff: male (in %)	-0.000
	(0.003)
Staff: pedagogical degree (in %)	-0.001
	(0.001)
Staff: fulltime (in %)	0.006***
	(0.002)
Citizens per km2	-0.001
	(0.001)
Employed female (in %)	0.003
	(0.005)
Fulltime-Employed female (in % of total working)	-0.000
	(0.001)
Share 0-6-y-old children (in %)	0.034
	(0.025)
Votes for CDU and FDP (in %)	-0.020**
	(0.009)
Votes for other parties (in %)	-0.007
	(0.007)
Log of GDPpc (in 1000Euro/Citizen)	0.079
	(0.130)
Local business tax rate	-0.120
	(0.157)
Local tax rate on agrarian real property	0.006
	(0.073)
Local tax rate on other real property	-0.045
	(0.060)
Municipality FE	Yes
Reference year/cohort	2004
Joint significance - Slots (p-Value)	0.765
Joint significance - Institution (p-Value)	0.887
Joint significance - Center (p-Value)	0.001
Joint significance - Regional (p-Value)	0.043
Adj. R2	0.939
Children	93570

Notes: This table shows the coefficients of the OLS estimates of the share of full-day slots. Regressions control further for a full set of municipality dummies and a constant term. Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.
Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord. Own calculations.

Table 3.4: Effects of the Share of Full-Day Slots on Child Development

	School Readiness	Motor skills	Socioemotional maturity
Share of fullday slots	0.058	0.039	-0.222***
	(0.043)	(0.066)	(0.078)
Municipality Fixed Effects	Yes	Yes	Yes
Cohort Fixed Effects	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
Regional Controls	Yes	Yes	Yes
Quality Controls	Yes	Yes	Yes
Adj. R2	0.097	0.083	0.091
Children	93570	93570	93570

Notes: This table displays the estimates resulting from an OLS regression of children's development at school entrance on the share of full-day slots measured in children's year of birth. In column 1, the school readiness indicator (dummy equal to one if the child is assessed to be ready to follow the school curriculum) is regressed on the share of full-day slots (continuous measure from 0-1) and on the full set of individual, regional and care center characteristics as well as a set of municipality and cohort dummies. In column 2, the motor skills indicator (dummy equal to one if the child has no motoric skill problems) and in column 3, the socio-emotional development indicator (dummy equal to one if the child is assessed to be socio-emotionally mature) are the dependent variables of the same specification, respectively. Standard errors are clustered at the municipality level.
Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord, Own Calculations.

diploma – Abitur – or not), and parental cohabitation status (single versus two parent household). Some interesting heterogeneities arise.

Stratifying the results by gender does not reveal any significant differences (see Table 3.5, Panel A). The expansion of full-day slots by 11 ppt leads to a deterioration in socio-emotional maturity among boys and girls (2.7 ppt vs. 2.1 ppt, respectively). Among girls, we observe a significant gain in terms of school readiness (0.08 ppt). Among boys, we observe a similar, but insignificant gain in terms of overall school readiness (0.06).

Differences are much more striking in terms of children's ethnic background (see Table 3.5, Panel B): immigrant children suffer much more in terms of their socio-emotional development than native children. While native children score on average 2.3 ppt lower when full-day care increases by 11 ppt, immigrant children score on average 3.9 ppt lower. This difference is significant at the 5% significance level. There is another notable difference in the effect of full-day care between native and immigrant children, which is highly policy relevant: immigrant children benefit substantially from the expansion of full-day care in terms of how prepared they are for primary school: the observed expansion leads to an increase in immigrant children's overall school readiness by 2.6 ppt. This improvement is substantial in light of the raw gap between immigrant and native children of 7.6 ppt in terms of school readiness (87.9% of all native children are assessed to be ready for school, while only 80.3% of all immigrant children are).

Similarly striking are the results when stratifying by parental education (see Table 3.5, Panel C): while children from a more educated family background are not affected by the observed expansion in full-day slots, children from a less educated family background experience substantial losses in terms of their socio-emotional development (2.7 ppt).

The strongest detrimental effects in terms of socio-emotional maturity arise among children living with one parent only (see Table 3.5, Panel D): while children from two-parent households lose "only" 2.4 ppt in the assessment of their socio-emotional

Table 3.5: Strata: Effects of the Share of Full-Day Slots on Child Development

	School Readiness	Motor skills	Socioemotional maturity
Panel A: Gender			
Boys			
Share of full-day slots	0.050	0.034	-0.244***
	(0.063)	(0.095)	(0.091)
Adj. R2	0.097	0.059	0.086
Children	49016	49016	49016
Girls			
Share of full-day slots	0.070**	0.044	-0.195***
	(0.032)	(0.044)	(0.072)
Adj. R2	0.075	0.040	0.076
Children	44554	44554	44554
F-test (p-value)	0.695	0.890	0.358
Panel B: Immigrant ancestry			
Native			
Share of full-day slots	0.030	0.027	-0.206**
	(0.040)	(0.065)	(0.078)
Adj. R2	0.089	0.088	0.092
Children	77330	77330	77330
Immigrant			
Share of full-day slots	0.232**	0.073	-0.355***
	(0.089)	(0.089)	(0.103)
Adj. R2	0.134	0.068	0.082
Children	11690	11690	11690
F-test (p-value)	0.011	0.551	0.046
Panel C: Parents' education			
High Education			
Share of full-day slots	0.032	0.070	-0.114
	(0.039)	(0.081)	(0.082)
Adj. R2	0.050	0.087	0.074
Children	17864	17864	17864
Low Education			
Share of full-day slots	0.065	0.038	-0.244***
	(0.050)	(0.069)	(0.080)
Adj. R2	0.096	0.081	0.093
Children	74856	74856	74856
F-test (p-value)	0.475	0.518	0.036
Panel D: Household composition			
Cohabiting Parents			
Share of full-day slots	0.062	0.015	-0.215***
	(0.046)	(0.066)	(0.080)
Adj. R2	0.097	0.083	0.085
Children	73569	73569	73569
Single Parents			
Share of full-day slots	0.034	0.162	-0.347***
	(0.079)	(0.105)	(0.107)
Adj. R2	0.099	0.088	0.101
Children	14102	14102	14102
F-test (p-value)	0.693	0.070	0.046

Notes: This table displays the estimates resulting from an OLS regression of the three respective children's development indicators (dummies equal to one if there is no development problem) at school entrance on the share of full-day slots measured in children's year of birth and on the full set of individual, regional and care center characteristics as well as a set of municipality and cohort dummies. The four panels show the regression results for different sample subgroups. Standard errors are clustered at the municipality level.

Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord. Own Calculations.

maturity, children residing with just one parent lose 3.8 ppt when full-day care increases by 11 ppt.

3.5.2 Robustness

Results based on our baseline specification rest upon several assumptions which we address and test in turn. Table 3.6 shows the results of the respective alternative specifications. First, we assume independence on the share of full-day slots conditional on a series of municipality and care center features. Our baseline specification conditions on the share of full-day slots and municipality and care center features measured in the same year (the year of child birth). In an alternative specification, we opt for conditioning on the municipality and care center features in the year prior to child birth and thus one year prior to measuring the share of full-day slots (see Table 3.6, Panel A). Doing so shall account for the lag in expanding child care slots after predicting the necessary expansion. Second, as pointed out in Section 3.3.1, data on full-day slots and care center features refer to all children enrolled in care centers. As such, these features can not be disaggregated by age groups. Our baseline specification controls for the share of full-day slots in the year of child birth. This shall rule out any endogeneity of the available full-day slots to parents' choice of enrolling their children in day care on a full-day basis. In an alternative specification, we use the average share of available full-day slots across the years when a child is potentially enrolled in day care and thus when a child is between 0-6 years old (see Table 3.6, Panel B). Finally, our baseline specification rests upon the assumption that there are no region-specific changes over time that are orthogonal to the municipality and care center characteristics we already control for and that affect children belonging to one cohort and living within a given municipality differentially. While it is hard to come up with any systematic difference within a municipality over time that is likely to drive the expansion of full-day slots and affects children's development beyond the conditions we already account for, we estimate an alternative specification where we allow for region-specific time trends (see Table 3.6, Panel C).

Table 3.6 shows the results of the three alternative specifications for the most affected subgroups: immigrant children, children from low educated backgrounds and children living with one parent only. The results for immigrant children are remarkably robust: the loss in immigrant children's socio-emotional maturity is comparable in magnitude and significant across all three alternative specifications, the gain in school readiness is also comparable in magnitude across all three alternative specifications, but only renders significance in two out of three specifications. The results for children from low educated backgrounds are robust to choosing a differential timing for the full-day slots and the municipality and child care features (see Panel A), but lose magnitude and precision when opting for an alternative time window of the full-day slots (see Panel B) and controlling for region-specific time trends (see Panel C). Finally, the results for children from a single parent household are robust across all alternative specifications, with the exception of controlling for region-specific time trends (see Panel C): in this case, the coefficient loses not only precision, but also magnitude.

3.6 Potential Counterfactual Care and Mechanisms

The estimation results – based on the SEE data – show that on average longer opening hours of child care centers do not affect children's cognitive skills or motor skills, but harm their socio-emotional development. Yet, as stated earlier, these effects are to be interpreted as intention-to-treat effects because our data do not provide us with any information whether children attend child care on a full-day or on a half-day basis. Consequently, we cannot rule out that these effects are driven by endogenous selection into full-day care, by the share of compliers, or by differential compliance across subgroups.

In order to gain some intuition on the mechanisms driving the results, we draw on an additional dataset called "*Familie in Deutschland*" (FiD). This dataset is provided by the *Deutsches Institut für Wirtschaftsforschung* (DIW) and part of the well-established German Socio-Economic Panel (SOEP). The FiD data are based on family surveys and

Table 3.6: Robustness Checks: Effects of the Share of Full-Day Slots on Child Development

	School Readiness	Motor skills	Socioemotional maturity
Panel A: Lagged Determinants of Full-day slots			
Immigrant Ancestry			
Share of full-day slots	0.248***	0.054	-0.345***
	(0.079)	(0.092)	(0.127)
Low Education			
Share of full-day slots	0.086*	0.058	-0.236***
	(0.048)	(0.071)	(0.088)
Single Parent			
Share of full-day slots	0.091	0.201*	-0.331***
	(0.071)	(0.105)	(0.107)
Panel B: Full-day slots at age 0-6 years			
Immigrant Ancestry			
Share of full-day slots	0.180	0.009	-0.581***
	(0.162)	(0.253)	(0.202)
Low Education			
Share of full-day slots	0.097	-0.069	-0.170
	(0.068)	(0.132)	(0.127)
Single Parent			
Share of full-day slots	0.166	0.177	-0.413**
	(0.117)	(0.224)	(0.165)
Panel C: Region-specific time trends			
Immigrant Ancestry			
Share of full-day slots	0.186	-0.112	-0.252*
	(0.125)	(0.091)	(0.148)
Low Education			
Share of full-day slots	0.017	-0.098	-0.098
	(0.051)	(0.067)	(0.083)
Single Parent			
Share of full-day slots	0.082	0.046	-0.128
	(0.095)	(0.133)	(0.118)

Notes: This table displays the estimates resulting from an OLS regression of the three respective children's development indicators at school entrance on the share of full-day slots using three alternative specifications. In panel A, all regional and care center characteristics are measured one year prior the year that the share of full-day slots is measured. In panel B, the average share of available full-day slots across the years when a child is born and the year when the child turns 6 years old is built, and in the same vein the average of the center and regional characteristics is built. In panel C, a regional time trend is added to the baseline specification. The estimation results for three different subgroups are shown. Standard errors are clustered at the municipality level.
Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord. Own Calculations.

include birth cohorts from 2004-2007, i.e. we have an overlay from 2004-2006 with the birth cohorts covered by the SEE data. Families are interviewed when their children are 4-5 years old, i.e. they are on average one year younger than children in the SEE. The FiD data do not provide information on school readiness or motor skills, but evaluate socio-emotional ability by help of the SDQ questionnaire – as done in the SEE.[20] Unlike the SEE data, the FiD data cover the whole of Germany, not only one state. We restrict the analysis to West Germany and end up with 1,103 children living in 251 counties. The main advantage of the FiD data is the included information on children's actual full-day care attendance. Thus, via the county identifier we are able to link data on regional and center characteristics with data on individual care attendance provided by the FiD data.

3.6.1 Potential Counterfactual Care

We do not know the counterfactual care mode for children in full-day care. The provided information in the FiD dataset, however, helps to assess the role of maternal care and alternative care modes for children in full-day care in contrast to those in half-day care. Panel A of Table 3.7 shows the average number of hours of care which are not spent with the mother. It indicates that the main difference in terms of care mode between children in full-day care and children in half-day care is the number of hours of care provided by the center instead of being provided by the mother. Children in full-day care are spending 16 hours per week more in a care center than children in half-day care. Time spent in other care modes – provided by other family members or external private care – does not significantly differ between the two groups of children.[21] This is an important contrast to evaluations of child care in countries with more private market oriented child care systems: In Germany, almost zero substitution towards private or informal care is taking place since private care is very expensive in Germany. There are quite generous tax schemes and a labor market legislation that makes it (financially) more attractive to mothers of young children to

[20] Table 3.10 in the appendix shows descriptive statistics of the child characteristics in the FiD data and compares them with those in the SEE data.

[21] Table 3.11 in the appendix shows that this is also the case when stratifying the sample.

stay at home than to put their child in expensive private care. In sum, given the child care infrastructure in Germany it is very unlikely that the counterfactual care would be non-maternal care in absence of full-day care slots.

3.6.2 Mechanisms

Overall effects

Positive effects of attending a care center on a full-day basis could accrue due to spending the major part of the day under the guidance of staff working in the care center. Care centers have the clear educational mission to develop children's motor, language, and pattern-recognition skills. Staff members support the development of these skills by engaging children in playful activities. Moreover, parents may benefit from feedback provided by center staff; in other words, there may be positive spill-over effects on parents' child caring practice. Table 3.7, Panel B provides evidence on the activities parents undertake with their children on a daily basis. There is, however, no evidence on positive spill-over effects on parents' child caring practices. If anything we see that parents whose children are enrolled in child care on a full-day basis are likely to pursue fewer outdoor activities and manual activities such as drawing or doing some arts and crafts. Also the income effect of using full-day care could have consequences for children's development. Given the reduction of maternal care, enrolling children in full-day care is likely to help mothers to engage in employment. Table 3.7, Panel C provides descriptive statistics on maternal employment: mothers whose children are enrolled in childcare on a full-day basis are 18.7 ppt more likely to be employed, 9.5 ppt more likely to work on a full-time basis, which translates in an increase of 7.3 working hours per week. In terms of available household net income, this means an increase of 288.6 Euro/month, which corresponds to an increase of 48.4 % in terms of the average household net income among families whose children are enrolled on a half-day basis. This increase in families' net income, could partly explain positive effects on children's skills (Dahl and Lochner, 2012; Gonzalez, 2013).

Table 3.7: Descriptive Evidence on Potential Channels

	Full-day	Half-day	Unconditional difference	Conditional difference
Panel A: Care modes				
Hours: Care center	35.121	18.715	16.406***	15.884***
			(0.474)	(0.518)
Hours: Partner	13.8	13.427	0.373	2.458*
			(1.280)	(1.408)
Hours: Grandparents	3.288	3.029	0.259	0.506
			(0.396)	(0.403)
Hours: Siblings	0.701	0.935	-0.233	-0.066
			(0.205)	(0.219)
Hours: Relatives	0.401	0.213	0.187*	0.155
			(0.104)	(0.107)
Hours: Others	0.468	0.35	0.118	0.071
			(0.089)	(0.088)
Panel B: Activities with child				
Going to playground	0.378	0.36	0.018	0.002
			(0.032)	(0.033)
Actions outside (walks o. similar)	0.808	0.873	-.065***	-0.070***
			(0.022)	(0.024)
Visiting oth. families w. children	0.35	0.366	-.017	-0.016
			(0.032)	(0.036)
Going shopping w. child	0.45	0.422	0.028	0.024
			(0.032)	(0.037)
Singing children's songs	0.233	0.267	-0.034	-0.007
			(0.027)	(0.029)
Drawing or doing arts/crafts	0.493	0.591	-0.098***	-0.083**
			(0.034)	(0.035)
Card games or similar	0.499	0.534	-0.035	-0.053
			(0.031)	(0.034)
Watching TV/videos/DVD w. child	0.601	0.583	0.018	0.020
			(0.029)	(0.032)
Playing computer game w. child	0.108	0.114	-0.006	-0.005
			(0.019)	(0.020)
Reading/telling stories	0.814	0.818	-0.004	-0.014
			(0.025)	(0.027)
Panel C: Maternal employment				
Employed	0.73	0.544	0.185***	0.187***
			(0.031)	(0.033)
Fulltime-Employed	0.233	0.137	0.096***	0.095***
			(0.026)	(0.028)
Working hours	20.174	12.697	7.477***	7.287***
			(1.027)	(1.117)
Labour Net Income	954.398	595.89	358.508**	288.6***
			(81.646)	(78.72)

Notes: This table shows evidence from an alternative dataset, the "*Familie in Deutschland*"-survey conducted by the *DIW Berlin*. The data stem from parents' interviews, conducted between the years 2010-2013, on 1103 children living in 251 counties in West Germany who were born between 2004-2007. Panel A shows the weekly hours a child is taken care of by someone other than the mother. Panel B displays activities the mother is undertaking with her child (dummy equals one if an activity is undertaken at least several times a week). Panel C shows the labor market participation of the mother. Labour net income is measured in Euro/month and working hours in hours/week. In column 1 and 2 mean values for children in full-day care and half-day care are shown respectively. In column 3 the raw difference between the two groups is calculated. In column 4, the variables of interest are regressed on full-day attendance controlling for individual characteristics, regional characteristics, quality characteristics of child care settings as well as for state and year effects. Standard errors in columns 3 and 4 are clustered at the county level and shown in parentheses: * p<0.1, ** p<0.05, *** p<0.01.

Source: *Familie in Deutschland*. Statistics of the *Child and Youth Services in Germany*. INKAR. Statistik-Nord. Own calculations.

Importantly, there is likely to be heterogeneity in the returns to substituting time spent with the mother by time spent in a care center across the different skill dimensions: while center staff is well trained in stimulating the development of cognitive skills, center staff may be less successful in supporting the development of children's emotional skills. Attachment theory postulates that the separation from the primary caregiver – who in most cases is the mother – causes anxiety and stress of children. In other words, the attachment theory establishes that time spent with the mother is crucial for the development of children's emotional skills (Bowlby, 1969; Mercer, 2006). As a consequence, this theory could explain why we find negative effects of full-day care – leading to a reducation of valuable time spent with the mother – on children's socio-emotional maturity.

As we find heterogeneous effects of full-day care across subgroups, however, we now try to understand the mechanisms behind this heterogeneity by help of the FiD data.

Effect heterogeneity

If the counterfactual care mode of full-day care is maternal care, the effects of expanding the opening hours of care centers depend on the returns to the time (not) spent with the mother. Obviously, there is a great deal of heterogeneity in these returns. A well-established literature in psychology and sociology on educational disparities in parenting styles stresses, for instance, that children have more to gain (and more to lose) from (not) spending time with better-educated mothers, see Hsin and Felfe (2014). For example, better educated parents are more verbally engaged with their children (Hart and Risley, 1995), provide more cognitive stimulation at home (Davis-Kean, 2005; Linver *et al.*, 2002; Yeung *et al.*, 2002), and have higher academic expectations for their children (Davis-Kean, 2005). In contrast, the time children spend with less-educated mothers may be more conflicting. The stress induced by socioeconomic disadvantage is known to create harsh and inconsistent parenting (McLoyd, 1998). Less-educated parents are also more likely to hold jobs that accrue lower prestige, offer lower pay and fewer benefits, are more unstable, and expose workers to greater physical hazards and psychological stress, all of which are

factors that are known to negatively correlate with child development (Felfe and Hsin, 2012; Han, 2005; Johnson *et al.*, 2005; Raver, 2003).

One explanation of the different effects of full-day care found in Section 3.5 could be a different take-up rate across sub-groups of children who benefit differently from full-day care. Column 1 in Table 3.8 shows the results of first stage regressions using the FiD data. Childrens's actual full-day care attendance is regressed on children's county's full-day share measured at children's year of birth. The regressions include county fixed effects as well as time fixed effects and regional and individual characteristics. While the overall take-up rate is at 0.33 – i.e. when three full-day slots are established one full-day slot is taken – the take-up rate is highest for children of immigrant ancestry (84% of new slots would be taken). This could be one explanation for our finding in Section 3.5 that the expansion of full-day care has a larger impact on the skills of immigrant children.

Furthermore, heterogeneous effects found in Section 3.5 could also be driven by different income effects across subgroups when full-day care enables mothers to work (again or for longer hours). If high-educated mothers have a higher income gain when sending their child to full-day care they are more able to provide material compensation for the reduction of time spent with their child. Additionally, mothers with a larger income gain are eventually less stressed because they use the additional income to outsource household chores. Results of some further regressions with the FiD data give empirical support to this potential mechanism. In Table 3.8, Column 2-4, we regress maternal labour market outcomes on children's full-day care attendance and include county fixed effects as well as time fixed effects and regional and individual characteristics. Column 2 shows that mothers of all subgroups whose child is attending full-day care are (by 15-19 ppt) more likely to be employed – with the exception of single mothers for whom there is no employment effect. While the effect on the intensity of labor market participation is about the same for all subgroups (except single mothers), namely between 7 and 8 hours more per week, the income effect varies across subgroups. Low-educated mothers and foreign mothers have the lowest income gain (about 250 Euro more per month), whereas high-educated mothers have the largest income gain (about 410 Euro more per month). Given this evidence,

our finding that children of high-educated mothers – in contrast to peers with more disadvantaged backgrounds – are not negatively affected by attending full-day care, could plausibly be due to more financial resources of their mothers to compensate for the lack of maternal time.

A further explanation of why we find heterogeneous effects of expanding the full-day share could be a different quality of care depending on the socio-economic composition of a municipality. As stated before we only possess structural quality measures provided by the statistics of the *Child and Youth Services in Germany* to proxy the actual quality in child care centers. To get some idea whether child care quality varies across regions depending on the socio-economic composition we show in Table 3.12 in the appendix the average of quality indicators in the lowest and highest quartile of the respective subgroups (high-educated parents, single parents, immigrant parents). The table shows that there are two groups of children who are more likely to be provided with a better structural quality: Children of high-educated parents as well as children of immigrant ancestry are more likely to be in smaller playgroups (19 children vs. 23 children per group resp. 20 vs. 22 children per group) and have more educated staff (67% vs. 60% with pedagogical degree resp. 63% vs. 60% with pedagogical degree). Thus, there is some evidence for different quality of care across subgroups, yet, it is hard to believe that the different findings across subgroups found in Section 3.5 could be completely due to these quality differences.

It is more likely that a combination of the mechanisms described in this section is responsible for the heterogeneous effects of full-day care: 1) Time in full-day care implies for all children a reduction of time with their mother. High-educated mothers, however, are more able to compensate for the lack of time with their child due to a higher income effect and thus, we do not find a negative effect of full-day care on high-educated mother's children. 2) When employees in child care centers possess better skills to train some of children's skills than parents do, as in the case for immigrant children and their language skills, longer hours in care can foment children's overall school readiness. 3) Children of different subgroups react differently to a higher provision of full-day slots. Children of immigrant ancestry have the highest take-up rate. This could explain why we find strongest effects of full-day care attendance

Table 3.8: Fullday Care Take up and Maternal Employment by Subgroup - Descriptive Evidence

	Take-up	Employed	Working hours	Net Income
Panel A				
All	0.333**	0.186***	7.298***	290.1***
N=1103	(0.152)	(0.033)	(1.108)	(77.74)
Panel B				
High-educated Parents	0.436	0.164**	6.978***	409.3*
N=299	(0.363)	(0.070)	(2.396)	(227.9)
Low-educated Parents	0.289	0.183***	7.373***	246.7***
N=797	(0.208)	(0.038)	(1.214)	(57.76)
Panel C				
Foreign Parents	0.840**	0.156*	3.540	256.8*
N=201	(0.344)	(0.0804)	(2.791)	(143.7)
Native Parents	0.250	0.189***	7.907***	295.9***
N=902	(0.164)	(0.038)	(1.295)	(90.47)
Panel D				
Single Parents	0.466	-0.006	0.822	-66.95
N=113	(0.558)	(0.105)	(3.093)	(168.4)
Cohabiting Parents	0.374**	0.193***	7.427***	317.6***
N=990	(0.164)	(0.035)	(1.171)	(81.51)

Notes: This table shows evidence from an alternative dataset, the "*Familie in Deutschland*"-survey conducted by the *DIW Berlin*. The data stem from parents' interviews, conducted between the years 2010-2013, on 1103 children living in 251 counties in West Germany who were born between 2004-2007. Column 1 shows the results of first stage regressions using the FiD data. Childrens's actual full-day care attendance is regressed on children's county's full-day share measured at the children's year of birth. The regressions include county fixed effects as well as time fixed effects and regional and individual characteristics. In Column 2-4, we regress maternal labour market outcomes on children's full-day care attendance and include county fixed effects as well as time fixed effects and regional and individual characteristics. In Panel A, the whole sample is used for the analysis, panel B, C and D separately analyze the subgroups.
Standard errors are clustered at the county level and shown in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.
Source: *Familie in Deutschland*. Statistics of the *Child and Youth Services in Germany*. INKAR. Statistik-Nord. Own calculations.

for this subgroup. 4) Regions with a higher share of high-educated parents and regions with a higher share of immigrants have on average a better structural child care quality. This could partly explain why we find positive effects of full-day care on immigrant children's school readiness and no detrimental effects on the socio-emotional maturity of children with high-educated parents.

3.7 Conclusion

What are the implications for children when expanding the supply of full-day slots in care centers? We analyze this question using two particularly rich administrative data sources for one West German state – school administrative data on children's development at the onset of primary school and care center records on the supply and quality of center-based care. Our identification strategy relies on several reforms triggering the expansion of full-day slots in public child care. Specifically, we exploit the municipality-specific deviations from the overall trend in the timing of the expansion.

Our results show that on average longer opening hours of child care centers do not affect children's cognitive skills or motor skills, but harm children's socio-emotional development. Alternatively stated, spending less time with the mother has a negative impact on children's socio-emotional maturity. Subgroup analysis reveals that these effects are driven by children from disadvantaged backgrounds, in particularly, children of immigrant ancestry, from low educated families and single parent households. Children from more advantaged backgrounds are not affected at all, which is most likely explained by their parents being more successful in compensating for the loss of time during the afternoon. Importantly, subgroup analysis also reveals that children of immigrant ancestry are better prepared for primary school – a finding which is most likely explained by center staff being better prepared to train immigrant children's language skills than their parents.

Our findings are interesting from two points of view: first, they highlight once again that the consequences of center-based care depend on the quality of the alternative care modes. Second, they make clear that the returns to time investments are likely to be non-linear and thus, one cannot just extrapolate from studies analyzing, for instance, the effects of center-based care functioning on a half-day basis.

Appendix

Table 3.10: Comparison of SEE and FiD Data

	SEE data		FiD data		
	Pooled (N=93,570)	SD	Full-day (N=489)	Half-day (N=614)	Full- vs. Half-day
Socio-Emotional Maturity(D)	.825	.380	.863	.889	-.026
Age (in month)	73.761	3.888	52.877	59.349	-6.471***
Male (D)	.524	.499	.54	.502	.038
Immigrants (D)	.125	.331	.176	.187	-.011
Single parent (D)	.151	.358	.139	.073	.066***
Mom's education: high (D)	.274	.446	.333	.221	.112***

Notes: Individual characteristics in column 1 and 2 are based on the full sample (2004-2012) of the SEE data. Column 3 and 4 display the individual characteristics for children in the *Familie in Deutschland*-survey; the means for children in full-day care and half-day care are shown respectively. In column 5, the raw difference between column 3 and column 4 is calculated.

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

Source: School entrance examination 2004-2012 and *"Familie in Deutschland"*-Survey. Own calculations.

Table 3.9: Determinants of the Share of Fullday Slots Using Lagged Municipality Features

	Full-Day-Share at Age 1
Cohort 2005 (D)	0.014*
	(0.008)
Cohort 2006 (D)	0.027*
	(0.015)
Cohort 2007 (D)	0.046**
	(0.021)
Cohort 2008 (D)	0.094***
	(0.025)
Cohort 2009 (D)	0.113***
	(0.026)
Cohort 2010 (D)	0.130***
	(0.027)
Cohort 2011 (D)	0.205***
	(0.048)
Cohort 2012 (D)	0.255***
	(0.053)
Coverage 0-3 years old (in %)	0.003
	(0.006)
Coverage 3-6 years old (in %)	-0.002
	(0.003)
Public provider (in %)	-0.000
	(0.001)
Other provider (in %)	-0.000
	(0.001)
Children per group	-0.002
	(0.003)
Age of staff (years)	-0.003
	(0.005)
Staff: male (in %)	-0.002
	(0.004)
Staff: pedagogical degree (in %)	-0.001
	(0.001)
Staff: fulltime (in %)	0.005***
	(0.001)
Citizens per km2	-0.001
	(0.001)
Employed female (in %)	0.006
	(0.004)
Fulltime-Employed female (in % of total working)	0.001
	(0.001)
Share 0-6-y-old children (in %)	0.024
	(0.021)
Votes for CDU and FDP (in %)	-0.017**
	(0.008)
Votes for other parties (in %)	-0.009
	(0.006)
Log of GDPpc (in 1000Euro/Citizen)	0.007
	(0.103)
Local business tax rate	-0.096
	(0.146)
Local tax rate on agrarian real property	-0.036
	(0.065)
Local tax rate on other real property	0.035
	(0.045)
Municipality FE	Yes
Reference year/cohort	2004
Joint significance - Slots (p-Value)	0.774
Joint significance - Institution (p-Value)	0.866
Joint significance - Center (p-Value)	0.001
Joint significance - Regional (p-Value)	0.213
Adj. R2	0.944
Children	93570

Notes: This table shows the coefficients of the OLS estimates of the share of fullday slots when children are one year old. All regional characteristics are measured at children's year of birth. Regressions control further for a full set of municipality dummies. Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.

Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR/Statistik-Nord, Own calculations

Table 3.11: Hours in Fullday Care and Alternative Care Modes - Descriptive Evidence

	Hours Care center	Hours Partner	Hours Grandparents	Hours Siblings	Hours Relatives	Hours Others
Panel A						
All	15.884***	2.458*	.506	-.067	.156	.072
N=1103	(.519)	(1.408)	(.403)	(.2196)	(.1076)	(.0886)
Panel B						
High-educated Parents	13.908***	3.523	.016	-.511	-.049	.079
N=299	(.740)	(3.375)	(.721)	(.604)	(.092)	(.196)
Low-educated Parents	16.688***	2.667	.719	.136	.228*	.124
N=797	(.633)	(1.669)	(.529)	(.210)	(.137)	(.090)
Panel C						
Foreign Parents	15.714***	-.730	.518	.785	.120	-.123
N=201	(1.135)	(3.022)	(.814)	(1.040)	(.299)	(.126)
Native Parents	15.842***	3.140**	.340	-.287	.151	.083
N=902	(.534)	(1.587)	(.443)	(.176)	(.125)	(.099)
Panel D						
Single Parents	18.454***	.021	-1.259	-0.126	.028	.573
N=113	(1.683)	(.081)	(1.725)	(0.301)	(.401)	(.457)
Cohabiting Parents	15.324***	3.072*	.629	-.028	.101	.027
N=990	(.521)	(1.606)	(.416)	(.247)	(.106)	(.092)

Notes: This table shows evidence from an alternative dataset, the "*Familie in Deutschland*"-survey conducted by the *DIW Berlin*. The data stem from parents' interviews, conducted between the years 2010-2013, on 1103 children living in 251 counties in West Germany who were born between 2004-2007. We regress (weekly) hours spent in center-based care and alternative care modes on children's full-day care attendance and include county fixed effects as well as time fixed effects and regional and individual characteristics. In Panel A, the whole sample is used for the analysis, panel B, C and D separately analyze the subgroups.
Source: *Familie in Deutschland*. Statistics of the *Child and Youth Services in Germany*. INKAR. Statistik-Nord. Own calculations.

Table 3.12: Care Quality and Demographic Characteristics - Descriptive Evidence

	(1) Low education (<p25)	(2) High education (>p75)	(3) Difference (2)-(1)	(4) Few single parents (<p25)	(5) Many single parents (>p75)	(6) Difference (5)-(4)	(7) Few immigrants (<p25)	(8) Many immigrants (>p75)	(9) Difference (8)-(7)
Demographics									
Share of high-educated parents	0.099	0.397	0.298***	0.329	0.259	-0.070***	0.281	0.275	-0.005
			(0.009)			(0.014)			(0.015)
Share of single parents	0.125	0.104	-0.021***	0.048	0.199	0.151***	0.092	0.151	0.060***
			(0.007)			(0.004)			(0.009)
Share of foreign parents	0.06	0.055	-0.005	0.034	0.129	0.095***	0.006	0.181	0.175***
			(0.007)			(0.007)			(0.004)
Care Quantity at Birth									
Coverage 0-3 years old (in %)	2.673	4.416	1.743***	3.378	3.685	0.307	3.153	4.204	1.051***
			(0.198)			(0.231)			(0.213)
Coverage 3-6 years old (in %)	79.471	82.834	3.364***	81.098	80.63	-0.468	80.995	82.642	1.647***
			(0.502)			(0.653)			(0.558)
Full-Day Share	0.129	0.238	0.109***	0.121	0.288	0.167***	0.093	0.325	0.232***
			(0.017)			(0.019)			(0.016)
Care Quality at Birth									
Public provider(in %)	22.513	17.463	-5.049*	18.705	18.915	0.210	22.233	17.06	-5.173**
			(2.634)			(2.451)			(2.338)
Other provider (in %)	24.444	41.093	16.649***	29.21	36.591	7.381***	22.443	37.959	15.516***
			(2.525)			(2.735)			(2.542)
Church provider(in %)	53.043	41.443	-11.600***	52.085	44.494	-7.590***	55.324	44.981	-10.344***
			(2.644)			(2.685)			(2.575)
Children per group	23.312	18.983	-4.329***	21.19	20.692	-0.497	22.201	20.19	-2.012***
			(0.470)			(0.483)			(0.457)
Age of staff (years)	38.827	39.133	0.306	39.323	38.526	-0.797**	39.621	38.553	-1.068***
			(0.326)			(0.311)			(0.314)
Staff: male (in %)	1.846	3.475	1.629***	2.128	3.95	1.821***	1.986	4.256	2.269***
			(0.408)						(0.413)
Staff: pedagogical degree (in %)	60.284	66.813	6.529***	61.608	61.55	-0.058	60.797	63.161	2.365**
			(1.270)			(1.418)			(1.149)
Staff: fulltime (in %)	19.01	27.908	8.898***	16.98	34.649	17.669***	14.707	37.815	23.109***
			(1.854)			(1.752)			(1.407)

Notes: This table displays the averages of care quantity and care quality in the lowest and highest quartile of the respective subgroup (high-educated parents, single parents, immigrant parents). Colums 3, 6 and 9 display the difference between the quartile's averages.
Standard errors are shown in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.
Source: Statistics of the *Child and Youth Services in Germany*/School entrance examination 2004-2012/INKAR. Own Calculations.

Chapter 4

Regulating Child Care Markets – Center-based Care vs. Family Day Care in Germany

4.1 Introduction

Public child care has been a prominent political and social topic in the last two decades. Subsidized universal care for the youngest is not only seen as an effective way to positively influence the development of all children, but also to foster maternal labor force participation.

While there is an increasing literature on the effect of child care on female employment and on its effect on children's educational outcomes (Berlinski *et al.*, 2009; Gormley Jr. *et al.*, 2008; Cascio, 2009; Datta-Gupta and Simonsen, 2010; Havnes and Mogstad, 2011; Noboa Hidalgo and Urzua, 2012) there is not much evidence yet on how political decisions – via regulations – shape the quantity and quality of early childhood institutions and by that the child care market (one of the few studies is Hotz and Xiao (2011)). Since the literature agrees that care quality matters for potential benefits children gain

from early child care it is important to understand how legislation can affect quality measures.

In this paper, I investigate the link between the political goal to increase the supply of early child care on one hand, and the stringency of quality regulations on the other hand. How does the relation between the two affect the child care market? And, moreover, how do the resulting market characteristics affect child development?

For studying these questions, I use a federal reform in Germany that triggered a large expansion in early child care. In 2008, federal law stated that children younger than 3 years old would be entitled to a slot in child care from 2013 on (while 3-6-year-old children had already been entitled to a slot since the late 1990s). From then on, not only large investments in early child care centers were undertaken, but at the same time rules were established that allowed also slots in family day care homes to be regarded as a valid fulfillment of the legal entitlement. Both of these care modes are publicly subsidized and, importantly, parental fees are about the same size. At the same time, both care modes are very different in terms of regulation: While center-based care underlies stricter regulation with respect to staff education and facilities, family day carers only have to meet rather low educational standards and mostly provide care in their own homes.

In Germany, the provision of child care falls under the responsibility of municipalities and it was thus the task of local authorities to manage a care expansion sufficient to fulfill the legal claim. Meanwhile, formulating quality standards (staff-per-child-ratio, qualification of staff and annual training requirements) is done at the state level. In response to the federal law triggering the early care expansion the data reveal a large variation in how the different states reacted being confronted with a quantity-quality trade-off (e.g. lowering or increasing the staff-per-child-ratio).

In the first part of my analysis, I exploit this variation of state regulations in combination with unique administrative panel data on the universe of child care centers and on all family day care homes in Germany. In addition, I have access to detailed information on structural quality indicators of both care modes for the years 2006-

2011. I aggregate and merge this data at the county level with data on the regional socio-demographic characteristics. Using 2300 county-year-observations and building on the empirical approach of Hotz and Xiao (2011) I investigate whether state-specific quality regulations had an impact on the care mode that German counties established. Using a specification including state fixed effects as well as a set of regional control variables avoids biased results due to policy endogeneity. My findings suggest that family day care homes were established predominantly in counties with a lower GDP per capita, less living space per inhabitant, and more female employment. This shows that counties react to economic incentives that the expansion of family day care homes offers: for the local authorities, family day care is less expensive to establish than center-based care. Family day care therefore offers a suitable way for poorer regions to fulfill the federal requirement of increasing the supply of early child care slots according to demand. Stricter state regulations targeted at the quality of this care mode, however, can act as a disincentive to expansion: I find that especially in poorer regions, the average supply reaction to stricter staff qualification regulations is negative.

In the second part of my analysis, I look at the implications of an increase in a county's share of family day care on a set of child development outcomes measured by the school entrance examinations of one large German state. In contrast to previous literature and opposed to public opinion, results indicate that a larger share of family day care is not detrimental for child development. Children seem to benefit in terms of their socio-economic maturity and motor skills. Including quality measures of family day care in the regressions, however, also matters. A positive effect on children's socio-emotional maturity is only found when quality indicators such as group size and quality of staff are controlled for. This part of my analysis provides evidence that important features of non-parental care at an early age include small group sizes and a more family-like context. Consequently, family day care might indeed represent an efficient way to harmonize two policy targets when expanding early child care in a fast manner: not to put too much of a burden on public finance and not to harm children's wellbeing.

The study proceeds as follows: Section 4.2 presents an overview of previous literature on regulations in the child care market and the importance of child care quality. Section 4.3 provides information on the data used for my analysis and describes characteristics of the German child care system. Section 4.4 explains the empirical approach to estimate the effects of regulations on family day care and presents the results. Section 4.5 describes the empirical approach to estimate the effect of the family day care share on child development and presents the results. Section 4.6 concludes.

4.2 Literature

There is a growing literature in economics on the effects of child care. One strand of the economic literature deals with the effects of child care supply on female employment, i.e. asks whether the policy goal to increase the female labor market participation is achieved by providing more child care. Another strand of the economic literature is studying the effects of child care attendance on children's skills and later labor market outcomes. This literature finds that human capital is built from the very early years on and that, consequently, providing child care is a means to increase human capital in the whole economy if child care has positive effects on children (Cunha *et al.*, 2006; Blau and Currie, 2006).

Recent studies investigating the impact of an expansion of universal child care on children's skills differ in their findings. Felfe *et al.* (2015), Cascio (2009) and Havnes and Mogstad (2011) find positive effects on children's skills, Baker *et al.* (2008) find negative effects. Potential reasons for these differences are different methodological approaches or different educational systems, but may partly also be explained by differences in quality of care across the respective regions of study.[1]

[1] How is child care quality measured? The cleanest approach to measure quality would be to measure how staff working in child care is interacting with children, which skill-stimulating actions are undertaken, how the day is structured. It turns out, however, that it is very hard to get objective information on such process quality indicators. Therefore, existing literature proxies child care quality by structural quality parameters such as group size, staff-to-child ratio and staff training (see Blau and Currie (2006)). One of the earliest studies on the question how to measure quality of child care is the one by Blau (1997).

Therefore, it is important to pay attention to the role of care quality when investigating potential benefits of child care on maternal labor market participation and children's skills. A recent study on the effects of care quality on maternal employment decisions finds that quality of child care indeed matters (Schober and Spieß, 2014). According to the authors, a lower staff-to-child ratio is negatively linked to maternal employment, i.e. it seems that mothers are more reluctant to use out-of-family care if the quality is low.[2] In line with the findings of Schober and Spieß (2014) the study of Simonsen (2005) finds that the provision of high-quality child care has a positive effect on mothers' employment. Yet, she also documents a negative price effect: the higher the cost (i.e. the lower the subsidies) of high-quality child care the lower maternal employment. Also Blau and Hagy (1998) who study determinants of the demand for child care quality find that lower-priced care modes have a higher demand. They report that a decreasing price of care leads to a decreasing demand for quality of that care.

As regards effects of child care quality on child outcomes existing studies agree that quality matters. Investigating this question in a situation where no structural quality parameters are at hand is feasible when differentiating between different child care modes (that have different quality features). One recent study dealing with the care quality question in this manner is the one by Datta-Gupta and Simonsen (2015). They differentiate between center-based care and family-day care in Denmark. Both care modes are regulated by the Danish state in terms of educational content, safety and hygiene. Yet, care center staff must hold a pedagogical degree while family day carers only have to attend shorter vocational courses. In contrast to more formal center-based care, family day care is taking place at the home of a carer who minds a group of up to 5 children. The authors claim that center-based care dominates family day care in many quality dimensions. Using an instrumental variable approach they find that children benefit more from attending center-based care than from attending family

[2] Interestingly, even though maternal employment decisions are affected by the observable staff-to-child ratio they are not affected by other quality indicators like staff education. The authors conclude that only easily observable quality characteristics matter for mothers' decisions for or against using child care arrangements.

day care.[3] The authors suggest that their findings are best explained by the different educational background of staff in the two modes of care.[4] Similar to these findings are those of an earlier study by Datta-Gupta and Simonsen (2010) in which the authors look at short-run effects of attending center-based care versus family day care using the SDQ average[5] of 5-year-old children as outcome of interest. Also for this age group they find that being enrolled in center-based care relative to family day care is beneficial for children.[6] The findings of Datta-Gupta and Simonsen (2015, 2010) are in line with Bernal and Keane (2011) who find that attendance of informal care reduces cognitive skills in contrast to attendance of center-based care.[7] They estimate that an additional year of informal care causes a 2.6% reduction in test scores whereas formal center-based care has no adverse effect on cognitive achievement and suppose that the reason for their finding is twofold: First, center-based care providers have better educated staff who may provide more cognitive simulation to children than informal providers, and second, center-based care may enable more stimulating interaction with other children and more educational activity than informal care.

Given the important role of child care quality for reaching both main targets of child care policies – better child outcomes and higher maternal labor market participation – there is surprisingly little evidence yet on how political decisions influence the quality of early childhood care modes. Blau (2007) stresses that there could emerge

---

3 Datta-Gupta and Simonsen (2015) exploit variation across municipalities in guaranteed access to center based care as an instrument for care mode and their results show that center based day care improves grades in Danish language in the final year of compulsory school with around 0.2 standard deviations.

4 They argue that the characteristics of family day carers are similar to those of low educated mothers while staff in child care centers is more similar to the home environment of high educated parents providing more verbal interaction with the child.

5 The *Strengths and Difficulties Questionnaire*, short SDQ, (Goodman, 1997) is a questionnaire designed to identify emotional problems, behavioral problems, hyperactivity, peer relationships, and social behavior.

6 Especially male children of low educated mothers are positively affected by center-based care attendance. Interestingly, female children of high educated mothers are negatively affected by attending center-based care instead of family-day care. Apparently, there can be heterogeneous effects of care quality. Especially girls may profit from the smaller group and family-like atmosphere in family-day care settings.

7 The definition of informal care by Johnson *et al.* (2015) is : "care provided on a regular basis to children from birth through age 5 by unlicensed, noncustodial caregivers. Other terms for informal child care are family, friend, and neighbor care; home-based care; kith and kin care; relative care; and license-exempt care." Therefore, family day care in the US can be seen as a subcategory of informal care.

unintended consequences of child care regulations.[8] This is in particular important in the context of policy targets aiming at increasing the quantity of child care: Does increasing the supply of child care by imposing regulations such as a legal entitlement for a child care slot come along with a decrease in child care quality? Apparently, this was the case in Canada. Baker *et al.* (2008) find negative effects of an early child care expansion and stress that the slot creation was done too fast. As a consequence, the quality of care could not catch up simultaneously.

Hotz and Xiao (2011) study the consequences of US state regulations on the child care market. They estimate the effect of the introduction of stricter quality rules on the number of accredited child care centers and on the emergence of unregulated care (which is family-day care in the US case). For identification they use changes in the regulations of child care quality (minimum staff-to-child ratio, group size and staff qualification) exploiting panel data between 1987 and 1997. Hotz and Xiao (2011) show that stricter rules lead to a reduction of the number of accredited child care centers and to an increase in the number of children taken care of in unregulated family day care. As regards the remaining child care centers, the regulations indeed lead to an improvement in quality. Importantly, a reduction of the number of child-care centers rather took place in low income areas, while quality improvements took place in high income areas. Thus, the authors show that enforcing quality regulations worsen the situation for families and children in low income areas (by reducing the number of centers and worsening the conditions in family-day care), but improve the conditions in high income areas.

In the first part of this paper, I analyze the consequences of child care regulations for the provided care mode in a county. Following a federal reform aimed at increasing the supply of early child care the German states respectively reacted with quality

[8] Blau (2007) justifies the need for having regulations in the child care sector as follows: The need for regulation is given because child care markets are markets with asymmetric information. Since parents are ex ante not well informed about the actual quality of a provider and often do not have the means to assess the quality, the government should guarantee some minimum level of quality. Bastos and Cristia (2012), for example, find that – in the absence of regulations – low care quality comes along with low income of consumers: They use census data on child care providers from Sao Paulo, Brazil, and document that particularly in poor districts a high share of child care centers does not reach the recommended quality level. In richer districts, however, the share of private care suppliers is higher and their quality is increasing parallel to household income in these districts.

regulations varying across states and in their timing. In the same way as Hotz and Xiao (2011), I use state and year fixed effects as well as regional control variables to mitigate the policy endogeneity bias problem when estimating the effects of regulations on the provided child care mode. In contrast to Hotz and Xiao (2011) I look at the effects of regulations in a country where private, non-regulated care is very uncommon so that there is no substitution between regulated and non-regulated care taking place when stricter regulations are imposed. In contrast to the US, the two child care modes in Germany – center-based care vs. family day care – differ in the stringency of regulations, but are both subject to quality audits of local authorities. A further difference to Hotz and Xiao (2011) is the use of the family day care share (i.e. provided slots in family day care out of all slots provided in public child care) as dependent variable instead of using the absolute number of family day carers as dependent variable.

In the second part of this paper, I look at the impact of a higher family day care share on children's development. In contrast to Datta-Gupta and Simonsen (2010, 2015) I rely on a fixed-effect approach instead of relying on an instrumental variable approach to estimate this impact. Furthermore, my data only allow to estimate an intention-to-treat effect. Yet, opposed to their studies, I am able to include quality parameters such as group size and staff qualification in my analysis. This allows me to discuss the specific role of quality inputs in child care whereas Datta-Gupta and Simonsen (2010, 2015) can only speculate on which aspect of family day care is responsible for the detrimental effect on child development that they find.

Before presenting my empirical approach, the following section gives an overview of the data and the child care system in Germany.

4.3 Data and Institutional Framework

4.3.1 Data

For my analysis of child care regulations in Germany, I exploit administrative data on all child care institutions in Germany for the years 2006-2011. The two main data sets I use are part of the Statistics of the *Child and Youth Services* in Germany (*Kinder- und Jugendhilfestatistik*). The data cover the universe of child care centers and all family day care homes in Germany. Not only information on the number of children and staff in each center is included, but also detailed information on structural quality indicators, such as group size, education of staff and staff-to-child ratio. Unfortunately, the individual centers and family day carers cannot be followed across years since the individual center identifier changes yearly. Due to this restriction and since data protection rules also prevent use of municipality averages, I rely on county averages of child care parameters in order to create a panel data set containing 411 county observations per year.

Via the county identifier I merge this panel of administrative child care data with regional socio-demographic characteristics (GDP per capita, unemployment, female employment, share of women working part-time, population density, living space per inhabitant, share of 0-3-year-olds in the population, migration inflows and outflows). These data are included in the so-called *INKAR-Raumordnungsdaten*, a data set on county characteristics published by the *Federal Institute for Research on Building, Urban Affairs and Spatial Development* (BBSR). I end up with with a panel consisting of 2,299 county-year-observations.

In order to investigate the implications of the expansion of family day care for children's development in the second part of my analysis, I use another administrative data set providing information about children's school readiness measured at the

School Entrance Examination (SEE)[9] in one big German state, namely North Rhine-Westphalia.[10] SEE data contain medical assessments on, among other dimensions, children's language skills, motor skills and socio-emotional maturity and can take five forms: "normal development", "some problems, but no treatment is necessary", "some problems, already in treatment", "problems, treatment necessary", and "problems which will reduce the child's performance in school". Based on these categories, I construct – in the same manner as in Chapter 3 of this thesis – binary indicators respectively for language and motor skills as well as socio-emotional maturity. The indicators are equal to one if the child does not exhibit any problem in the assessed tasks. The SEE data also contain the pediatrician's assessment whether a child is ready to follow the school curriculum or not. The recommendation can take the following three forms: "ready for school", "school enrollment only with support provided by the teacher", and "special needs education required". I construct – again exactly as in Chapter 3 of this thesis – a binary indicator which equals one if the child is assessed ready for school. Additionally, I use obesity (a binary variable equal to 1 if a child is obese) as health outcome in my analysis. The SEE data from North Rhine-Westphalia have the advantage that they include almost 100,000 children every year who live in 53 different counties, i.e. there is sufficient statistical power to employ a county fixed effect approach. Via the county identifier I merge the SEE data with the two other data sets presented above. Variation of center-based care and family day care between counties and across time can then be used to measure the implications for children's skills at school entrance age. The SEE data, however, do not include any direct information on the care mode a child attended (center-based care or family day care). As in Chapter 3 of this thesis, I therefore can only provide estimates for the intention-to-treat effect (ITT), but not for the treatment effect itself.

[9] School Entrance Examination data are also used by Felfe and Lalive (2014), Cornelissen *et al.* (2016), and in Chapter 3 of this thesis; all of these studies, however, are focusing on SEE data from different German states.

[10] The SEE dataset is subject to major restrictions concerning data protection legislation. Due to these regulations I had to analyze the data via on-site use at the *Landeszentrum Gesundheit Nordrhein-Westfalen*, part of the Ministry of Health, Equalities, Care and Ageing in North Rhine-Westphalia.

4.3.2 Organization of Child Care in Germany

Child care in Germany is organized at the municipality and county level. In contrast to the US, child care is predominantly publicly provided. Less than 4% of all children attend a for-profit center (Schilling, 2009). Children are taken care of in three different types of center-based care depending on their age: Until the age of 3 they are attending early child care centers (*Krippen*). Children between 3 and 6 years are attending kindergarten and school-aged children are taken care of in after-school care (see Chapter 2 of this thesis on effects of after-school care on child development and Chapter 3 on effects of full-day kindergarten on child development). Alternatively, children can attend family-day care which mostly takes place at the private homes of family-carers and is mainly targeted at children between 0 and 3 years.

Compared to other OECD countries, parental fees for child care are very low in Germany, on average they lie between 61 and 161 Euro per month depending on age group and on hours of care (Müller *et al.*, 2013). Importantly, in contrast to other more private-market oriented child care systems, parents are in general not able to get their child a higher-quality child care experience by paying more money. This is due to a tight regulation of parental fees in public day care (Schober and Spieß, 2014).

According to data of the OECD (2012), the expenditures on early child care amounted to a bit more than 0.6% of GDP in Germany in 2009. This is slightly above the OECD average. Denmark, in contrast, spends more than 1% of GDP on early child care. That means, the studies by Datta-Gupta and Simonsen (2010, 2015) are presenting evidence on a child care system allocating significantly more resources on child care than the German system does. On the other hand, in the US, only 0.42% of GDP are paid on child care. Consequently, the studies by Blau (2007) and Hotz and Xiao (2011) are referring to a system with significantly fewer resources spent on child care. These different financial situations are mirrored in some of the structural quality indicators such as the staff-to-child ratio: In Denmark, the average kindergarten staff-per-child ratio is 1:7, in the US 1:12, and in Germany 1:10.

Table 4.1: Coverage for 0-3-Year-Olds over Time 2007-2011

Year	2007	2009	2011	2007-2011 Change
All Public Child Care Coverage - Mean	0.162	0.209	0.264	0.63
All Public Child Care Coverage - p25	0.066	0.117	0.166	1.51
All Public Child Care Coverage - p75	0.177	0.230	0.303	0.72
Family Day Care Coverage - Mean	0.020	0.028	0.036	0.85
Family Day Care Coverage - p25	0.005	0.010	0.015	2.34
Family Day Care Coverage - p75	0.023	0.033	0.047	1.09
Share of Family Day Care - Mean	0.140	0.152	0.154	0.10
Share of Family Day Care - p25	0.043	0.062	0.066	0.51
Share of Family Day Care - p75	0.207	0.220	0.216	0.04

Notes: *Family Day Care Coverage* is defined as share of 0-3-year-olds attending family day care out of the whole population in this age group. *All Public Day Care Coverage* is defined as share of 0-3-year-olds attending any kind of public child care out of the whole population in this age group. *Share of Family Day Care* is defined as as share of 0-3-year-olds attending family day care out of all 0-3-year-olds attending any kind of public child care.
Source: Statistics of the *Child and Youth Services in Germany*. Own Calculations.

4.3.2.1 Center-based Care and Family-Day Care

Early child care is provided in two different modes: Care centers and family-day care.[11] As Table 4.1 shows, most 0-3-year-olds enrolled in early child care are attending center-based care. But, nonetheless, an increasing number of children is enrolled in family day care. In 2011 15% of all children aged 0-3 years in non-parental care were attending family-day care. While in 2007, on average 104 children were attending family day care per county, in 2011 already 192 children were using this care mode, which is an increase of 85 percent, see Tables 4.1 and 4.2.

Family day care differs from center-based child care in several ways. Average center-based care for 0-3-year-olds typically takes place in centers that accommodate about 40 children in total, split into smaller groups of about 13 children. Each center employs pedagogical employees, assistants and administrative employees. The average staff-

[11] Family day care is also offered to children older than 3 years, but this age group uses family day care less often. Only 0.9% of children aged 3-6 years are attending family day care (Strunz, 2011).

Table 4.2: Expansion of Family Day Care over Time 2007-2011

Year	Family Day Care Homes per county	Family Day Care Homes Total (nationwide)	Family Day Care Children per county	Family Day Care Children Total (nationwide)
2007	81	32987	104	42385
2009	94	38424	148	60446
2011	103	42468	192	79111

Notes: Family day care is defined as publicly subsidized child care which takes place at the homes of family day carers. Before being allowed to take children in their care, family day carers are screened and interviewed by municipality officials.
Source: Statistics of the *Child and Youth Services in Germany*. Own Calculations.

to-child ratio is 1:5.[12] In order to work as a group leader an employee must have acquired two years of theoretical training and at least two years of practice in a child care center. Salary for those with an accomplished degree in child care is about 2400 Euro (gross) per month (Sell and Kukula, 2011). There is not much variation of salary across centers and counties since the employees are paid according to standards of the German TvöD (national general accord of salaries). The employees are furthermore subject to social insurance contributions, statutory health and accident insurance as well as old-age provision. Additionally, the employees have the right to take a fixed amount of vacation days (around 30 days) and are entitled to full pay during sick days. The opening hours of centers vary; legally they are requested to offer opening hours according to demand and remain open for at least four hours five days per week. About half of the children are staying more than 7 hours in care (Lange *et al.*, 2007). While states (Bundesländer) are in charge of regulating the quality of center-based care in Germany, municipalities and counties are in charge of monitoring child care centers' compliance with these rules.[13]

In contrast to center-based care, family day care takes place in private homes of the family day carer. Again, the states are setting the rules concerning quality of this

[12] These are indicators for the structural quality of centers which are regulated by law. There are also legal rules concerning process quality in the form of pedagogical content. Every center has to have a pedagogical concept that must be approved by local authorities. There are no big scale data about process quality in Germany available yet. One first step is the study by Tietze *et al.* (2012).

[13] The legal requirements for center-based care for 0-3 year-olds did, importantly, not change since the major reform in 2008. Thus, there is only variation across states in terms of family day care regulations.

care mode and municipalities and counties are in charge of monitoring compliance. Depending on the state, there may be up to six children in each home (eventually including the carer's own children in the respective age group). The rules regarding facilities, security and hygiene are less clearly stated than in the case of center-based care (Schnock, 2009). In contrast to employees in center-based care family day carers are not required to have attained any degree in child care or pedagogy, but only have to get some training in short vocational courses with a minimum duration between 80-160 hours (depending on the state). Before being allowed to take children in their care, family day carers are screened and interviewed by district officials who assess the individual suitability and check the police clearance certificate of candidates (Schnock, 2009).

More than 90% of family day carers are self-employed (BMFSFJ, 2014). That means that they – in contrast to center-based staff – have to bear the complete cost of social insurance contribution, statutory health and accident insurance and old-age provision. In case of sickness, more than 50% of family day carers do not get paid for the missed days, i.e. they have a much higher risk of income loss than clerical employees in center-based care (Sell and Kukula, 2011). On top, family day carers are confronted with covering their operating costs.[14] In fact, family day carers are confronted with a hybrid job situation: They are self-employed in the sense that they have to take entrepreneurial risk, but they are not allowed to act as entrepreneurs in the sense that they are not free to negotiate their prices and services with families.[15]

[14] The following example may help to illustrate a family day carer's income situation: She can earn a gross income as high as a group leader working in center-based care. For this, however, she has to care for six children because she gets about 400 Euro per child per month. This way, she would get gross earnings of 2400 Euro per month. Her net earnings, however, would be much lower since she has to deduct her insurance and operating costs. A rough estimate of the monthly net income of family day carers caring for 6 children would thus be 1200-1400 Euro vs. center-based staff who end up at 1600 Euro monthly. Of course, these numbers depend on a couple of assumptions regarding operating costs, hours of care etc. To my knowledge, there are no data available to systematically compare the net earnings of both care modes.

[15] Sell and Kukula (2011) present the results of a survey among a representative group of family day carers in Germany. One of their findings is that family day carers can only cover all their living costs when caring for at least 5 children. Family day carers who are only taking in 3 children are thus dependent on their partner's earnings (about 70% only care for 3 children or less). In line with this fact, 51% of respondents would prefer to be clerical employees of the municipality rather than being self-employed but at the same time being subject to directives by municipality officials regarding quality and prices of their services.

Table 4.3: Quality in Family Day Care and Center-Based Care in 2011

	Family Day Care	Center-based Care
Children-per-Staff	2.85	4.96
	(0.82)	(2.01)
Group Size	2.85	13.49
	(0.82)	(4.43)
% of staff with pedagogical degree (*Erzieherausbildung*)	19.04	63.74
	(6.47)	(18.67)
% of staff with pedagogical diploma (*Diplompädagoge*)	3.34	4.52
	(1.15)	(6.23)
% of staff with short-term education degree	11.59	17.63
	(3.23)	(15.38)
% of staff without degree	65.18	−−
	(6.90)	−−
% of staff with family day care qualification	79.68	−−
	(11.50)	−−
% of staff with family day care qualification >=160 hours	45.73	−−
	(22.81)	−−

Notes: Family day care is defined as publicly subsidized child care which takes place at the homes of family day carers. Before being allowed to take children in their care, family day carers are screened and interviewed by municipality officials. Center-based care is public child care located in centers monitored by municipality officials. While in most states family day carers have to attend a qualification course, there are no requirements on the attainment of a degree or diploma in child care. This explains why 65% of family day carers do not have a degree. 80% of family day carers have attended a qualification course which shows that some of the family day carers hold both, a degree and a qualification course certificate, while some hold a degree only.
N=411, standard deviation in parenthesis.
Source: Statistics of the *Child and Youth Services in Germany*. Own Calculations.

When comparing the educational level of staff in center-based care versus family day care, it turns out that child care centers have higher qualified staff who is much more likely to have a degree in child care or pedagogy compared to family day carers (see Table 4.3). Yet, with respect to the children-per-staff ratio, the situation in family day care is better (at 1:3 instead of 1:5 in center-based care, cp. Table 4.3). Group size, consequently, is also much smaller in family day care with on average 3 children per group instead of 13 children per group (and even more per center). Due to the small group size, the opening hours of family day care are often quite flexible and care on week-ends is also offered. How does the situation for parents confronted with these two modes of public child care to choose from look like? Do prices of care differ depending on the care mode? Interestingly, in contrast to the US, parental fees in care centers and family day care are about the same in Germany. Independent of the

care mode, parents pay on average between 61 and 161 Euro per month (Müller *et al.*, 2013).[16]

Since the fees for both care modes are about the same, parental demand should be driven by the perceived quality of care mode. Yet, it is not clear whether one of the care modes is actually of superior quality. While staff in center-based care is significantly better qualified, there is clearly more potential for individualized care in family day care because of the small group size.[17] Additionally, family day care offers more flexible opening hours and may therefore seem more attractive to families. Yet, the *perceived* quality of center-based care is apparently higher than that of family day care: According to a representative survey 77% of parents prefer center-based care, while only 11% prefer family day care (BMFSFJ, 2014).[18] This is in line with the study of Datta-Gupta and Simonsen (2010, 2015) who use for their identification that demand for family day care is lower despite of similar fees as center-based care. In the US, in contrast, family day care is an option for parents who cannot afford prices of center-based care as Hotz and Xiao (2011) show; the authors argue, however, that lower prices come along with lower quality since family day care is not regulated by the state.

In Germany, as Schober and Spieß (2014) point out, the quality of day care children are confronted with is not or only weakly dependent on their parents' ability to pay higher fees. The question I investigate in this paper is whether children's municipality's (resp. county's) ability to pay is linked with the quality of care they receive. While the costs for parents are the same independent of the care mode, the costs for the municipality are quite different for the following reasons: Maintaining a child care center involves large fixed costs, operating costs, staff salaries etc. – independent of the number of children attending the center. In the case of family day care, the

[16] In Germany, parents only have to bear a small proportion of the total costs of public child care. The biggest part is paid by the municipalities (about 47 %) and by the state (about 31 %). Providers, like churches and welfare organizations contribute about 5 % and parents about 14 % (Schober and Spieß, 2014).

[17] Tietze *et al.* (2012) did one of the first attempts to find large-scale evidence of the process quality of child care in Germany. In a (non-representative) evaluation of family day carers they find that family day care is apparently not of worse process quality than center-based care.

[18] 6% would like to use a combination of care modes and the rest of parents has no preferences for one care mode.

municipality only has to pay a lump sum for each child who is actually taken care of. The municipality does not bear any operating cost and the entrepreneurial risk lies at the family day carer. Therefore, for a given municipality, the average yearly total costs of family day care are lower than those of center-based care while the parental fees are the same (Sell and Kukula, 2011). Consequently, municipalities face incentives to create child care slots in family day care rather than in center-based care in situations of economic distress and a contemporaneous high demand for early child care that legally has to be satisfied.

4.3.2.2 Reforms in the Child Care Market

In Germany, children are legally entitled to a slot in child care from their first birthday on. There were three major reforms in the child care sector that led to this entitlement: The first reform took place in the 1990s. Since 1996 every child turning 3 years old is legally entitled to a slot in a child care center.[19] As a result of this policy, since the early 2000s more than 90% of children entering school have attended kindergarten for at least 2 years. The second major reform in 2005 (the so-called "Tagesbetreuungsausbaugesetz") aimed at providing public child care also for younger children (between 0 and 3 years) of whom at this point in time only a very small percentage had access to early child care. The objective to make early child care universally accessible was reinforced in a third major reform ("Kinderförderungsgesetz") in December 2008, when the federal government announced that all children aged 1 year and older would be legally entitled to a child care slot from August 2013 on. It was specified that until then at least 35% of all children younger than 3 years old should have access to early child care. One important change that came together with this reform was the upgrading of family day care. The federal government proclaimed that family day care would be henceforth considered to be legally equal to center-based care. Consequently, the legal entitlement would be fulfilled regardless whether a municipality would offer a child a slot in center-based care or in family day care. The

[19] This reform's effects on child development are investigated by recent work of Cornelissen *et al.* (2016) and its effects on female employment are analyzed by Bauernschuster and Schlotter (2015).

specific numerical target for each municipality was set such that 33% of all early child care slots should be slots in family day care.

In the following years, a strong expansion of the numbers of family day care took place. The first column of Table 4.2 shows that there was an expansion from on average 81 family day care homes per district in 2007 to 103 family day care homes in 2011 (which is an increase of almost 30 percent). In total, the number of family day care homes in Germany rose from around 33,000 to around 42,500. The number of children attending these homes almost doubled in the same period. In 2007, there were on average 104 children per district (around 42,000 in Germany) taken care of in family day care, while in 2011 already 192 children (around 79,000 in Germany) attended family day care. The increase of child care centers was not as large in the corresponding period. The average increase was only about 5 percent (in 2007 there were on average 110 child care centers per district while in 2011 there were on average 116, cp. Table 4.8 in the appendix).

A look at the coverage rates of public child care in Table 4.1 reveals that the increase of slots per 100 children has been sizable in both care modes since the reforms. The coverage of all public child care (i.e. both care modes combined) rose from 16% in 2007 to 26% in 2011, while the largest increase took place in the lowest quartile (in the 25th percentile the coverage rate rose from around 7% to around 17% which is an increase of 150 percent). The major part of coverage is provided by center-based care, only 2 children out of 100 were taken care of in family day care in 2007 (while 14 children out of 100 were in center-based care). Until 2011, however, the coverage of family day care almost doubled and reached 3.6%. In the highest quartile the coverage was already at almost 5% in 2011. As stated above, the reform in 2008 aimed at providing at least 33% of all early child care slots in family day care by 2013. This has not yet been achieved in 2011 as the bottom part of Table 4.1 shows. Between 2007 and 2011, the share of family day care rose from 14% to 15.4%. Yet, comparing the lowest and highest quartile indicates that especially regions with low usage of family day care until 2007 started to expanding this care mode (with an increase of about 50 percent). The comparison between lowest and highest quartile also shows that there is large variation between regions with respect to their family day care share: In 2011

only every 20th child attended family day care in the lowest quartile, in the highest quartile every 5th child was taken care of in this mode (out of all children attending public child care).

While the federal government set the quantitative goal for the expansion of early care, it was up to the 16 German *Bundesländer* (states) to implement strategies to reach this goal. As in the US, states regulate the staff-to-child ratio, required qualifications and training of the staff, building and equipment safety etc. Interestingly, the *Bundesländer* did not change the quality regulation of center-based care after the 2008 reform. Regarding family-day care, though, they implemented the new role of family day care in their regulations after the reform, yet these regulations varied across states as regards their content and timing. For this paper, I collected information of regulation changes in family day care of the 16 German *Bundesländer* individually for the time span under study (by screening state laws and provisions dealing with child care provision and child care quality). I then coded the information in order to build indicators comparable to those in Hotz and Xiao (2011). Table 4.4 summarizes the changes in regulation displaying the means of state quality indicators for 2007, 2009 and 2011. It shows that rules became stricter, i.e. most of the *Bundesländer* started to regulate family day care more tightly in light of its new status (being legally equivalent to center-based care). While in 2007, on average up to 8 children were allowed to be taken care of in family day care, this was restricted to a maximum of 6 children per carer in 2011 in almost all states. There has been a significant increase in the number of required hours of qualification. In 2007, an amount of 80 hours of qualification was sufficient and in the major part of states there was not even a rule concerning the training. In 2011, the required amount of qualification hours was already at almost 110 hours on average and only 6 states did still not have any rule on this quality dimension. Furthermore, the number of required training hours per year increased. In 2007, about 5 hours annual training were reinforced by law in only 5 states, while in 2011 almost 7 hours annual training were requested by already 9 states.

Did the actual quality of family-day care change according to new regulations? I use the Statistics of the *Child and Youth Services in Germany* to (at least partly) answer

Table 4.4: Family Day Care Regulations in Germany from 2007-2011

Year	2007	2009	2011
Mean: Staff-per-child ratio in law	0.13	0.17	0.17
Number of states without Staff-per-child ratio rule	4	1	1
Mean: Number of qualification hours in law	79.13	102.63	108.75
Number of states without qualification hours rule	11	7	6
Mean: Number of training hours in law	5.44	6.38	6.94
Number of states without training hours rule	11	8	7

Notes: These quality indicators were built by analyzing the regulation changes of public child care law for each German state. The table displays the average of each indicator as well as the number of states without any rule concerning this indicator. The indicators are weighted by the number of counties in each state.
Source: Regulation Codes of 16 German states. Own Calculations.

this question: Table 4.5 shows in its first row that the children-per-staff-ratio has on average always been below the legally allowed maximum. The number of children per family day carer has increased from 2.1 to 2.9. This comes together with a decrease of groups with less than 4 children (in 2007 82% of family day care groups consisted of less than 4 children, in 2011 only 67%) and an increase of larger groups (in 2007 around 17% of groups consisted of 4-6 children, until 2011 this share had risen to 27%, in 2007 less than 1% of family day care groups included more than 6 children, in 2011 this share was already at 7%.). Table 4.5 furthermore shows that staff in family day care got more qualified on average between 2007 and 2011. In 2007 only about 50% of family day carers had completed a family day care qualification course, in 2011 already 80% had completed such a course. Importantly, the share of those carers who attended a course of at least 160 hours – as recommended by the German Youth Institute – increased from 17% to 46% between 2007 and 2011. Table 4.5 shows furthermore that the average number of weekly hours children attend family day care stayed about the same in the respective time period.

In sum, descriptive statistics indicate that – following the reform on early child care in 2008 – family day care became more prevalent on average and its quality improved in terms of staff qualification. In terms of group size and children-per-staff ratio there was a slight deterioration, however.

Table 4.5: Quality of Family Day Care over Time 2007-2011

Year	**2007**	**2009**	**2011**
Children-per-Staff-Ratio	2.146	2.465	2.850
	(0.719)	(0.727)	(0.822)
D: Groupsize smaller than 4	0.819	0.763	0.672
	(0.192)	(0.190)	(0.214)
D: Groupsize between 4 and 6	0.174	0.206	0.266
	(0.191)	(0.184)	(0.209)
D: Groupsize larger than 6	0.007	0.032	0.062
	(0.022)	(0.046)	(0.084)
% of staff with any family day care qualification	54.05	69.31	79.68
	(21.65)	(17.13)	(11.50)
% of staff with family day care qualification >=160 hours	16.54	27.04	45.73
	(19.50)	(21.18)	(22.81)
% of staff with pedagogical degree (*Erzieherausbildung*)	21.92	20.44	19.04
	(9.67)	(7.98)	(6.47)
% of staff with pedagogical diploma (*Diplompädagoge*)	2.60	3.27	3.34
	(1.20)	(1.87)	(1.15)
% of staff with short-term education degree	13.10	11.94	11.59
	(4.98)	(3.07)	(3.23)
% of staff with no degree/diploma	61.42	63.36	65.18
	(10.67)	(8.01)	(6.90)
Weekly Hours of care per child	29.272	28.483	28.779
	(8.207)	(8.459)	(8.712)

Notes: This table shows the average quality characteristics of family day care establishments in Germany for 2007, 2009 and 2011. "D" indicates a binary variable. N=411, standard deviation in parenthesis.
Source: Statistics of the *Child and Youth Services in Germany*. Own Calculations.

4.4 The Effects of Quality Regulations in Family Day Care

4.4.1 Estimating the Effects of Regulations

Measuring the impact of regulating the quality of family day care on the provided family day care share might lead to biased results when only relying on variation across states for identification. The reason is that regulations might be endogenous. Suppose, for example, that there are states with a traditionally high number of voters who give high-quality child care top priority and other states with a low number of voters with such preferences. If voters vote their preferences, then the establishment of stricter quality standards might be more likely in the former than in the latter. In addition, local authorities in counties of the former group of states might be more motivated to make sure stricter quality rules are complied with. Failure to control for the existence of such state-specific differences will bias the estimated effect of stricter qualification requirements on the provided family day care share.

My data – covering the 16 German Bundesländer (states) and the years 2006-2011 – allow, on the one hand, to control for regional characteristics such as GDP per capita and female employment, and on the other hand, to control for state and year fixed effects to net out any time-constant state-specific characteristics that could impact the establishment of family day care and also to control for the overall time trend to expand early child care coverage.

The following estimation approach relies on Hotz and Xiao (2011): In a first step, I estimate the following baseline specification.

$$Y_{cst} = \alpha_0 + \alpha_1 REG_{st} + \alpha_2 NoREG_{st} + \alpha_3 X_{cst} + \delta_t + \gamma_s + \epsilon_{cst} \tag{4.1}$$

where Y_{cst} is the share of family day care, measured as share of all 0-3-year-olds in family day care out of all 0-3-year-olds in public child care, in county c belonging to state s in year t.[20] The share of family day care is defined as share of 0-3-year-olds attending family day care out of all 0-3-year-olds attending public child care (i.e. either family day care or center-based care, see bottom part of Table 4.1 for descriptives.)

REG_{st} is a vector of the family day care regulations in state s in year t. $NoREG_{st}$ is a vector of binary variables that are equal to 1 if the state s did not have a family day care regulation in year t and are equal to 0 otherwise.

X_{cst} is a vector of county-level population characteristics (citizens per square kilometer, the population share of 0-3-year-old children, shares of immigration and emigration, the overall share of children between 0 and 3 years attending public child care, the share of non-public child care providers, i.e. church and welfare organizations) and economic conditions (log of GDP per capita, living space per inhabitant, unemployment, female employment and the share of part-time working women).[21] The sign of coefficients reflecting a county's economics (proxied by the variables GDP per capita, living space per inhabitant, and unemployment) is especially interesting in light of cost advantages for the public providers when supplying a higher share of family day care instead of center-based care (as explained in Section 4.3.2). Finally, δ_t are year fixed effects and γ_s are state fixed effects. As alternative specification to assess whether the inclusion of state fixed effects changes results, also estimates of a regression without state fixed effects are shown below.

In a second step, I investigate the heterogeneity of regulation effects by the following equation:

[20] Note that I am – due to data protection rules – not able to use municipality averages but have to aggregate the data on the second lowest regional level, the county.

[21] Table 4.9 in the appendix shows the averages of all county-level socio-economic characteristics over time.

$$Y_{cst} = \alpha_0 + \alpha_1 REG_{st} + \alpha_2 NoREG_{st} + \alpha_3 REG_{st} \cdot C_{cst} + \alpha_4 NoREG_{st} \cdot C_{cst} + \alpha_5 X_{cst} + \delta_t + \gamma_s + \epsilon_{cst} \tag{4.2}$$

where C_{cst} is a specific characteristic of a county. In my analysis, I use a county's GDP per capita and respectively, a county's population density as such a characteristic for the following reasons:

By estimation of this equation the heterogeneity of regulation effects between wealthier and poorer areas (resp. between more and less densely populated areas) is analyzed. On one hand, the size and sign of coefficients of interactions between regulations and county's GDP per capita indicate whether wealthier regions react differently to more stringent quality rules than poorer regions. Parents living in wealthier regions might be more demanding regarding quality aspects of care – and may thus put more pressure on local authorities to expand family day care only in the presence of stricter quality rules. On the other hand, looking at interactions between regulations and population density is of interest since establishing new child care centers is more expensive in urban areas because of lack of space.[22] Therefore, regulations may be more easily complied with in less densely populated areas.

As shown in Hotz and Xiao (2011), the estimation equations 4.1 and 4.2 do not only allow to estimate the *marginal effect* of a reform indicating the effect of a marginal change of an existing regulation. They also allow to calculate the *average effect* of a regulation established in a state where there had not been a regulation so far. Importantly, the marginal effect and the average effect of a regulation need not to be of the same sign since they refer to two different aspects of a regulation. It could be that the introduction of a minimum requirement of staff qualification in a system with no existing rules on this standard has a positive effect on expanding care (thus,

[22] Furthermore, urban areas face usually more constraints on the staff supply side than rural areas (Hüsken and Riedel, 2012). This is due to higher rental prices and higher cost of living in cities which hinder child care staff to move to cities or to qualify as child care staff when living in a city. In a follow-up version of my paper I plan to use average rental prices as proxy variable for constraints on the supply side of child care.

the average effect would have a positive sign). At the same time, increasing the requirement of staff qualification by 20 hours, for example, could have a negative effect on care expansion (thus, the marginal effect would be negative).

The average effect relying on the specification of empirical model (4.1) is calculated as follows:

$$E(Y_1 - Y_0|REG = \tilde{R}) = \alpha_1\tilde{R} - \alpha_2\tilde{R} \tag{4.3}$$

where $\tilde{R}$ is the mean of regulation REG.

The average effect relying on the specification of the empirical model (4.2) is given by:

$$E(Y_1 - Y_0|REG = \tilde{R}) = \alpha_1\tilde{R} + \alpha_3\tilde{R}\tilde{C} - \alpha_2 - \alpha_4\tilde{C} \tag{4.4}$$

where $\tilde{R}$ is the mean of regulation REG and $\tilde{C}$ is the mean of county characteristic C .

In sum, relying on the estimation approach of Hotz and Xiao (2011) I identify the marginal and average effects of state regulations on the family day care share expansion through the inclusion of state fixed effects and year fixed effects. By that, any time-invariant differences across counties due to being placed in different states are held constant and any shocks that affected all German states in the same year are held constant.[23]

4.4.2 Results: Effects on the Family Day Care Share

Table 4.6 shows the main results. Column 1 presents the estimated coefficients obtained with the empirical model (4.1). Columns 2 and 3 present the estimated co-

[23] In case that there are state-specific changes over time that are correlated with the regulation variables and that affect the share of family day care a county provides, the estimates would still be biased. In order to rule out this possibility, an additional specification in a follow-up version of this paper will allow for state-specific time trends.

efficients following the estimation approach using interaction terms as in empirical model (4.2).

Overall, it shows that state regulations on the staff-to-child-ratio, staff qualification and training hours have significant effects on the counties' family day care share.[24]

The findings in Column 1 indicate that a stricter rule on the staff-to-child ratio exerts a significant positive marginal effect on a county's family day care share. Reducing the maximal number of children per staff by 1 leads to an increase of the family day care share by 1 percentage point (ppt).[25] The average effect of regulations on the staff-per-child ratio is also positive, but quite small and not significant.

As regards regulations concerning the minimum number of required qualification hours, Column 1 in Table 4.6 shows that increasing the required amount by 30 hours would lead to a 1.2 ppt increase in the family day care share. Yet, the average effect of this regulation is negative: Introducing a regulation on the staff-per-child ratio at the overall mean of this indicator leads to a 6.4 ppt decrease in the family day care share which is quite sizable given the family day care share average of around 15% (see Table 4.1).

Turning to the marginal effect of regulations on the number of required annual training hours, Column 1 in Table 4.6 indicates a negative marginal effect of stricter rules: If the number of required training hours increases by 2, for example, the family day care share decreases by almost 1 ppt. The average of this regulation is not significant.

Overall, results in Column 1 show that regulations indeed matter for the provided share of family day care in a county. Yet, as discussed in Section 4.3.2, not only the legal framework matters for a county's child care policy but also its economic

[24] The following interpretations of results rest upon assuming linearity of effects of quality regulations on the family day care share. Yet, it may make a difference whether a regulation increases the minimum amount of qualification hours from 80 to 100 hours or from 140 to 160 hours. In a follow-up version of this paper, non-linearities of effects will be tested by including quadratic terms of regulations' stringency in the empirical model.

[25] The calculation is as follows: Increasing the staff-per-child ratio from, e.g. 0.143 to 0.173 (i.e. an increase of 0.03), is equivalent to reducing the maximal number of children from 7 to 6. The effect of reducing the maximal number of children per staff by 1 would thus lead to an increase of the family day care share of approx. 1 percentage point ($0.03 * 0.3 = 0.009$).

Table 4.6: Marginal and Average Effects of Quality Regulations on the Family Day Care Share

	(1)	(2)	(3)
Staff-per-child ratio in law	0.300**	1.507***	0.204
	(0.114)	(0.425)	(0.129)
No Staff-per-child ratio rule[c]	0.038*	0.292***	0.024
	(0.022)	(0.098)	(0.026)
Staff-per-child ratio in law x GDPpc (log)		-0.361***	
		(0.128)	
No Staff-per-child ratio rule x GDPpc (log)		-0.076**	
		(0.029)	
Staff-per-child ratio in law x population density			0.025***
			(0.007)
No Staff-per-child ratio rule x population density			0.004***
			(0.001)
Avg. Effect of regulation: minimum staff-per child ratio[a]	0.005	0.003	0.003
Number of qualification hours[b] in law	0.004**	-0.005	0.006***
	(0.002)	(0.006)	(0.002)
No qualification hours rule[c]	0.067**	0.088	0.091***
	(0.028)	(0.080)	(0.028)
Number of qualification hours[b] in law x GDPpc (log)		0.003*	
		(0.001)	
No qualification hours rule x GDPpc (log)		-0.004	
		(0.021)	
Number of qualification hours[b] in law x population density			-0.0001*
			(0.00006)
No qualification hours rule x population density			-0.003***
			(0.001)
Avg. Effect of regulation: minimum qualification hours[a]	*-0.064*	*-0.071*	*-0.070*
Number of training hours in law	-0.004***	-0.012	-0.005***
	(0.001)	(0.007)	(0.001)
No training hours rule[c]	-0.032	-0.228*	-0.046**
	(0.020)	(0.117)	(0.019)
Number of training hours in law x GDPpc (log)		0.002	
		(0.002)	
No training hours rule x GDPpc (log)		0.060	
		(0.037)	
Number of training hours in law x population density			0.0001
			(0.0001)
No training hours rule x population density			0.001
			(0.002)
Avg. Effect of regulation: minimum training hours[a]	0.006	0.387	0.010
Regional Controls	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Adj. R2	0.477	0.479	0.478
Observations	2299	2299	2299

Notes: [a]The average effect of a regulation is calculated according to equations 4.3 and 4.4. It is the effect of introducing a new regulation on a specific quality indicator at the mean value of this indicator relative to having no regulation on this indicator. Estimates of the average effects in *italics* are statistically significant at the 5%-level. [b]The number of required qualification hours is divided by 10 for the purpose of readability. [c]This binary variable equals 1 if there is no regulation on this quality indicator (otherwise it equals 0). Standard errors are clustered at the state-year level and are shown in parenthesis:

* p< 0.10, ** p<0.05, *** p<0.01.

Source: Statistics of the *Child and Youth Services in Germany*/INKAR. Own Calculations.

situation. Table 4.11 in the appendix shows the entire set of estimated coefficients of the baseline specification of the empirical model (4.1). It indicates that counties with a higher GDP per capita have a lower family day care share, i.e. they have a lower share of the less expensive child care mode. Moreover, Table 4.11 in the appendix reveals that the family day care share is higher in counties with a higher population density, less living space per inhabitant, more female employment, less part-time female employment and a higher inflow of people.

In order to investigate the role of county characteristics and their interplay with state regulations, columns 2 and 3 of Table 4.6 show results of regressions including interaction terms of regulations and county's GDP per capita and, respectively, county's population density. As one can see, effects of state regulations are heterogeneous across counties. Column 2 shows that the interactions of county's GDP per capita and regulations on the staff-per-child ratio and on the required number of qualification hours have statistically significant coefficients. While a more stringent rule on the staff-per-child ratio has a smaller effect in wealthier counties, a more stringent rule on the number of qualification hours has a larger effect in wealthier counties – a puzzling finding that is discussed below. As regards the required amount of training hours, there is no significant difference across counties of different GDP per capita. Column 3 shows that more stringent rules on the staff-per-child ratio have larger effects in more densely populated counties. More stringent rules on the the number of qualification hours, however, have a smaller effect in more densely populated counties. Again, rules on the required amount of training hours have no significantly different effects across more or less densely populated counties.

4.4.3 Discussion: Effects on the Family Day Care Share

As results show, both the legal framework and the economic conditions play a role for a county's implementation of family day care. Overall, counties with a higher GDP per capita have a lower family day care share. This indicates that family day care – being the less expensive care mode – is indeed an option for poorer counties

to expand their early child care system. In the appendix, I show a different approach to estimate the importance of a county's economics: I divide all counties into a *poor group* (all counties whose GDP per capita is below median GDP per capita in 2006) and a *rich group* (the rest of counties). Figure 4.1 shows that poorer counties started with a much lower share of family day care in 2006, but expanded this care mode in the following years more strongly than counties of the rich group. I estimate the effect of belonging to the poor group with a simple Difference-in-Difference approach interacting the status *poor* with a dummy indicating the time span after the reform in 2008. Table 4.10 in the appendix shows that the results of this approach are in line with the estimated coefficients of GDP per capita in table 4.11. According to the Difference-in-Difference approach belonging to the poor group significantly increased the share of family day care by 1.9 ppt after the reform.

Overall, results of Table 4.6 in Column 1 show that more stringent rules on training hours have a negative effect on a county's family day care share. This could be due to the following: Offering training to family day carers creates costs for counties and, therefore, rules increasing the required amount of training hours might make it less attractive for local authorities to engage in the search for family day carers. At the same time, also family day carers bear additional costs when they have to attend to more training hours per year which makes it less attractive to work in family day care.

As regards regulations on the required amount of staff qualification, Table 4.6 shows a small positive marginal effect which is more pronounced in wealthier counties and less pronounced in more urban counties. Yet, the average effect of the regulation is negative and quite large. As discussed in Section 4.3.2 counties face several constraints when expanding their supply of early child care. One important constraint is the lack of qualified staff. This could be responsible for the finding that the introduction of a state regulation on staff qualification leads to a decrease of the family day care share at first (resulting in a negative average effect). Once such a rule is established, increasing the amount of required qualification hours might attract more parents to use family day care. This would explain the larger effect of this regulation's stringency in wealthier counties in which parents might be more demanding "consumers". Additionally, wealthier counties might be financially more able to fulfill the higher

qualification standard by offering family day care qualification courses to interested persons.

Finally, a tougher regulation on the staff-to-child ratio has a positive effect on the family day care share. This might seem surprising at first, but could be due to the following: Since the *de facto* average staff-to-child ratio was already at 1:3 and below from 2007-2011 (cp. Table 4.5) new regulations that (at the maximum) established a ratio of 1:5 should not have presented a big obstacle for increasing the share of family day care, in other words the new rule did not *bite*. On the contrary, family day carers who used to care for 3 children only were maybe encouraged by the legal staff-to child ratio of 1:5 to take care of more children once the new rule was established. Due to that, the overall family day care share would have increased, too. Table 4.5 underpins this view: Between 2007 and 2011 the number of family day carers with groups of less than 4 children decreased while the shares of groups of 4-5 and more than 6 children increased. Table 4.6, Column 2 displays smaller consequences of a more stringent regulation on the staff-to-child ratio in wealthier counties and Column 3 displays larger consequences in more densely populated (i.e. more urban) counties. In line with the explanation above this could follow from more financial resources being available in wealthier counties to establish center-based care instead of family day care and to keep existing family day care at a small group size. In more urban regions, however, lack of space might be responsible for not establishing center-based care and instead rely on an expansion of family-day care and, additionally, on an increase of family-day care's group size.

One major difference with respect to the findings of Hotz and Xiao (2011) is that the stringency of regulations exerts no negative, but positive effects on the family day care share in the case of staff qualification and staff-to-child ratio in Germany. Only in the case of more stringent regulations on the required staff training, I find a small negative effect on the family day care share. Meanwhile, Hotz and Xiao (2011) find that more stringent regulations of these indicators on average decrease the number of child care centers in the US. The difference between the findings might be due to a different *bite* of regulations. While the standards in the US were set comparatively high (with respect to the ex-ante status) the standards in Germany were easily achievable by

market participants, especially in the case of the staff-per-child ratio. Importantly, the dependent variable in my approach is measured as share while Hotz and Xiao (2011) use absolute numbers of child care centers resp. family day care homes as outcome. In other words, I observe the impact of quality regulations as a switch in the relative supply of family day care, while Hotz and Xiao (2011) observe a decrease in the absolute number of centers. This means that also the different definition of outcome variables may be responsible for the different findings.

The main common finding of both studies is the importance of the economic standing of regions on the supplied care mode. Hotz and Xiao (2011) show that stricter regulations substitute formal care by informal care predominantly in poor areas. I find that especially poor counties expand their early child care system by means of expanding their family day care share, a much less expensive way for a county to supply child care. Moreover, I find that rules on staff qualification have a larger positive impact in wealthier regions which is also in line with Hotz and Xiao (2011).

Further robustness checks of the findings presented in this section will include an investigation of potential non-linearities of regulation effects. Furthermore, in order to understand the role of a county's existing child care provision for its reaction to stricter quality regulations, I plan to include interaction terms of regulation indicators with the existing composition and extent of child care in the empirical model.

4.5 Family Day Care and Child Development

4.5.1 Estimating the Effects of Family Day Care on Child Development

In this part of my analysis, I investigate the consequences of a county's share of family day care on child development. Hotz and Xiao (2011) only assume that the reforms they investigate and that led to a crowding-out of center-based care by family day

care would lead to detrimental effects on children. In contrast, I aim at providing an empirical answer to the question whether family day care is indeed the care mode with less beneficial, or more detrimental, effects on child development when it is accompanied by regulations on its quality that are evaluated in the first part of my analysis. Yet, my approach has some major limitations: The estimation of the effects of the family day care share on child development is done using data on children's skills provided by *one* state only (North Rhine-Westphalia, see Section 4.3.1 for a description of the data). Thus, a replication of the empirical models (4.1) and (4.2) for this region of study is not feasible since state-fixed effects cannot be implemented. In order to find out whether North Rhine-Westphalia is a representative region for studying the effects of family day care, I investigate the differences in the development of family day care over time (see below, Section 4.5.2) and do not find major differences. Furthermore, I compare the estimation results for the whole of Germany resulting from a regression with state fixed effects and one without state fixed effects and contrast the latter to results from a regression using only data from North Rhine-Westphalia. Finally, in a follow-up version of this paper I will use nation-wide survey data to combine the empirical models (4.1) and (4.2) with child outcomes.

In contrast to Datta-Gupta and Simonsen (2010, 2015) I do not use an instrumental variable approach to identify the effects of family day care on child outcomes but make use of the expansion of family day care due to the major reform in 2008. Thus, identification relies on a fixed effect approach similar to the one used in Chapter 3 of this thesis and is specified as follows:

$$Y^s_{ict} = \beta F_{ct_{-3}} + \gamma C_{ct_{-3}} + \delta X_{ct_{-3}} + \eta Z_i + \mu_c + \psi_t + \epsilon_{ict} \tag{4.5}$$

where Y^s_{ict} denotes skill dimension s of child i residing in county c at school entrance age (6 years old) at time t. $F_{ct_{-3}}$ stands for the share of family day care offered to children 3 years prior to school entrance, i.e. around the time when they have completed early child care and are about to switch to kindergarten care. $C_{ct_{-3}}$ represents a set of family day care quality features measured 3 years prior to school entrance, $X_{ct_{-3}}$ a

set of county characteristics measured 3 years prior to school entrance, and Z_i a set of individual background characteristics. μ_c are county fixed effects and ψ_t are year fixed effects.

Since the SEE data of North Rhine-Westphalia (for a description see Section 4.3.1) do not include information on the specific care mode a child was attending I can only estimate intention-to-treat effects, i.e. investigate whether a higher share of family day care in a given county leads to a higher or lower share of children being assessed ready for school (socio-emotionally mature/physically mature) in this county. Due to including county fixed effects, I control for any time-constant characteristics of counties that may correlate with the timing of the family day care share expansion and with changes in children's development. By including year fixed effects, I also control for the overall trend to expand early child care.

The main identifying assumption of my approach is that $F_{ct_{-3}}$ is conditionally independent of the unobserved determinants of children's development Y^s_{idt}. As Duncan and Gibson-Davis (2006) point out, the omitted-variable problem arises if unobserved characteristics of the child or family background are correlated with the choice of child care quality as well as child outcomes. I control for children's individual and background variables usually defined as being important for their development: gender, immigrant background, parental education, number of siblings, and birth weight. Still, it could be that especially parents with higher or lower (unobservable) ability of parenting send their child to family day care instead of center-based care when early child care is expanding. I assume that parents in Germany are on average not able to choose either mode of child care because of the low supply of early child care being confronted with a high demand. As described in BMFSFJ (2014), in 2012 early child care coverage was at 28%, but 40% of parents stated that they would like to enroll their child in early care. Given that counties fulfill their legal entitlement independent of offering a slot in center-based care or family day care it seems plausible that parents are not given a choice between the care modes. In consequence, I assume that children in both care modes have on average similar unobservable characteristics. In order to test this assumption, in a follow-up version of this paper, I will use the FiD data set – representative survey data presented

in Chapter 3 of this thesis – which includes information on the actual care mode attendance of children.[26]

4.5.2 Effects of Family Day Care on Child Development

Using administrative data of the School Entrance Examinations in North Rhine-Westphalia – Germany's territorial state with the highest population density and consisting of 53 counties – I analyze the intention-to-treat effect of family day care on children's skills at age 6.

Is the region of North Rhine-Westphalia representative for the rest of Germany with respect to its counties' expansion of the family day care share? Table 4.15 in the appendix shows that the expansion between 2007 and 2011 was about the same as for the whole of Germany: The coverage of family day care increased from 2% to 4%. The family day care share increased from 20% to 24% (which is a slightly larger increase than observed for the whole of Germany). Table 4.13 in the appendix shows that – as observed for the whole of Germany (shown in Section 4.4) – also in North Rhine-Westphalia especially counties in more economic distress increased their family day care share by more.[27]

As regards the quality regulations of family day care in North Rhine-Westphalia in comparison to the average state's quality regulations there are some, but no major, differences.[28] In sum, the state of North Rhine-Westphalia thus provides a useful exemplary region to investigate the implications of the expansion of family day care in Germany. Yet, when interpreting the effects one has to bear in mind that North Rhine-Westphalia's expansion was a bit above average. Furthermore, while the quality

[26] Furthermore, I am working on an instrumental variable approach using pre-reform numbers of training locations for family day carers per county as instrument for the family-day care share.

[27] There is a significant negative relationship between living space per inhabitant and the family day care share and also a negative (yet not significant) relationship between GDP per capita and the family day care share.

[28] The regulation of the staff-to-child ratio is less strict than in the average state. Furthermore, there is no regulation on the number of training hours family day carers have to attend. The regulation on required staff qualification hours, however, is stricter than in the average state (91 hours vs. 86 hours) and there are fewer years without any rule on this ratio than in the average state (42% of time vs. 64% of time between 2007 and 2011).

regulations on the staff-to child ratio are below average, the quality regulations on staff qualification are above average. Note also that parental fees and employment schemes of family day carers are not different from the German average.[29]

Turning to the results, Table 4.7 displays the estimated coefficients of fixed effect regressions based on the empirical model (4.5) discussed in Section 4.5.1.[30]

Panel A in Table 4.7 shows that a higher county family day care share comes along with a higher county share of children being assessed ready to follow the school curriculum and having better motor skills. When a county's family day care share increases by 10 ppt, this results in a 1.4 ppt increase in children's average school readiness and in a 1.1 ppt increase in children's average motor skills. The overall coverage of early child care (i.e. the sum of slots in center-based care and family day care divided by the sum of 0-3-year-old children) does not have a significant effect on children's skills at school entrance. Importantly, Panel A displays estimates of regressions that do not contain controls for the quality of family day care.

As Panel B in Table 4.7 shows, controlling for the quality of family day care makes a difference: The estimated effect of the family day care share on school readiness is no longer significant. Yet, the estimated effect on children's socio-emotional stability gets more precise and significantly positive: When a county's family day care share increases by 10 ppt, then chidren's social-emotional maturity measure increases by 0.7 ppt. The estimated effect on motor skills is about the same size as in the regressions without quality controls (indicating a 1.2 ppt increase coming along with a family day care share expansion of 10 ppt). As regards estimates of the coefficient of the overall coverage of early child care there is no difference with respect to size and precision compared to the regressions without quality measures.

[29] Estimating the empirical model (4.1) for North Rhine-Westphalia would only make sense if state's time-invariant characteristics do not play a role for the effect of child care regulations. In order to find out whether estimating the empirical model (4.1) without state fixed effects renders similar results as presented in Column 1 in Table 4.6 I present results of a regression without state fixed effects in Table 4.12 in the appendix. It shows that the results are different and change signs when not controlling for state fixed effects. Consequently, from Column 3 in Table 4.12, displaying results of estimating the empirical model (4.1) for the state of North Rhine-Westphalia only, might not be learned anything.

[30] Table 4.14 in the appendix shows the descriptive statistics of the sample of children used for this analysis.

Table 4.7: Implications of an Increasing Family Day Care Share for Children's Development

	Adipositas	School	Language	Motor	SocioEmotional
PANEL A:					
No Quality Controls					
Share of Family Day Care	-0.0016	0.1392*	0.0353	0.1076*	0.0231
at Age 3	(0.0106)	(0.0692)	(0.0925)	(0.0601)	(0.0336)
Coverage All Provided Child Care	0.0114	-0.0046	-0.0886	-0.038	-0.0751
at Age 3	(0.0228)	(0.1519)	(0.151)	(0.1169)	(0.0842)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	409323	288619	380920	383050	233459
Adj. R2	0.012	0.0491	0.0485	0.046	0.0571
PANEL B:					
With Quality Controls					
Share of Family Day Care	0.0098	0.1176	0.0263	0.1188*	0.0664*
at Age 3	(0.013)	(0.0783)	(0.091)	(0.0594)	(0.0367)
Coverage All Provided Child Care	0.0182	0.0033	-0.0937	-0.0391	-0.0829
at Age 3	(0.0235)	(0.1284)	(0.1539)	(0.1028)	(0.0734)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	409323	288619	380920	383050	233459
Adj. R2	0.012	0.0497	0.0485	0.0463	0.0572

Notes: This table displays the estimates resulting from a fixed effects regression of 5 child development indicators at school entrance on the family day care share measured when children are 3 years old. Outcomes are measured as binary variables and equal 1 if the pediatrician does not detect a problem in the respective skill dimension (except in the case of adipositas: a value of 1 indicates a positive diagnosis of obesity). Regressions include regional and individual characteristics as control variables and county fixed effects as well as year fixed effects. Standard errors are clustered at the county level and are shown in parenthesis:
* $p < 0.10$, ** $p<0.05$, *** $p<0.01$.
Source: Statistics of the *Child and Youth Services in Germany* and the School Entrance Examinations in North Rhine-Westphalia. Own Calculations.

4.5.3 Discussion: Effects on Children's Development

The finding of positive effects of an increasing share of family day care on children's development is quite surprising given the results of previous studies discussed in Section 4.2. Datta-Gupta and Simonsen (2010, 2015) report negative effects of family day care on children's socio-emotional development as well as on their later academic outcomes. Furthermore, Hotz and Xiao (2011) implicitly assume that family day care is of inferior quality to center-based care. Additionally, parents in Germany seem to be ignorant of potential beneficial effects of family day care attendance given their favoring of center-based care over family day care (see Section 4.3.2: 77% of parents prefer center-based care, while only 11% prefer family day care).

The identification strategy – using county fixed effects – relies on within-county variation of the family day care share over time. Results indicate that a positive change in a county's family day care share leads to a positive change in child outcomes. Meanwhile, a change in the overall supply of early child care apparently does not have a significant effect on child outcomes. One explanation for this could be that the changes in regulations came along simultaneously with the expansion of the family day care share and that those are responsible for the positive impact on children. The differences between coefficients of the regressions with and without quality indicators (Panel A vs. Panel B in Table 4.7) support this view. As shown in Table 4.5, the quality of family day care changed between 2007 and 2011: Complying with state regulations, the average family day carer got more qualified. The qualification of a family day carer is potentially most important for fostering a child's cognitive development as preparation to follow the school curriculum (measured via the school readiness indicator). Since family day carer's qualification improved parallel to the expansion of family day care, this leads to a positive effect of family day care on school readiness when not controlling for staff qualification. In the same vein, a small group size might be especially important for a child's socio-emotional development. Since group size increased parallel to the expansion of family day care I only find a positive significant effect of the family day care expansion when controlling for quality features. As stated before, the estimated effects are intention-to-treat effects

since the actual family day care attendance is not observed in the SEE data. Therefore, selection into family day care could also be responsible for the positive effect of family day care on children's development if especially children from families with better parenting skills attended family day care, for example. In this case, however, it would be necessary to explain why including quality measures should then make a difference in effects' size and significance. This would imply that selection into family day care changes according to altered quality features of family day care. To investigate the problem of potential selection, a future analysis will be based on the *FiD dataset* (presented in Chapter 3 of this thesis) – a representative survey of children in Germany that includes information on actual family day care attendance.

In sum, evidence drawn from North Rhine-Westphalia's SEE data indicates that the reform on family day care in Germany was quite successful. The combination of both quantity and quality targets led to an increase in counties' family day care share and to a positive overall effect of the expansion on children's motor and socio-emotional skills.

Why do the findings of this study differ from the evaluation of family day care in Denmark? One potential explanation could be the use of different estimation strategies: While Datta-Gupta and Simonsen (2010, 2015) rely on instrumental variable approaches to estimate the local treatment effect, I use a fixed effect method to estimate the intention-to-treat effect on the respective cohorts of children that were affected by an exogenous policy change. Furthermore, while Datta-Gupta and Simonsen (2010, 2015) look at effects of a well-established early child care system in Denmark, I investigate effects in a system that is only at the beginning of providing early child care to all children. This suggests non-linearities of the effects of the family day care share which will be investigated in a follow-up version of this paper. Finally, while quality of family day care in Denmark is comparable to quality of family day care in Germany, center-based care in Denmark is of higher quality compared to those in Germany. I.e. family day care might indeed be the less preferred option in Denmark while this does not have to be the case in the German child care system.

As regards the size of the effects presented in Table 4.7, effects are relatively small. An average expansion of the family day care share of 10 ppt would result in an increase of children's skills of about 2-3% of a standard deviation (compare the descriptive statistics of child outcomes in Table 4.14). Yet, in light of previous findings of detrimental effects of family day care in other countries it may be considered as relieving result that the expansion of family day care in Germany did at least not lead to negative consequences for children's development.

4.6 Conclusion

In this paper, on the one hand, I analyze the effects of the economic and legal environment on publicly provided child care quantity and quality using administrative data covering all child care centers and family day carers in Germany. On the other hand, I analyze the effects of the resulting child care quantity and quality on child development outcomes using administrative data covering four school entrance cohorts of Germany's territorial state with the highest population density (North Rhine-Westphalia) combined with data on provided slots in child care and child care quality.

In the first part of my analysis, I find that both the legal framework and the economic conditions play a role for a county's implementation of family day care. Overall, poorer counties establish a higher share of family day care. I furthermore find that rules on staff qualification have a stronger positive impact in wealthier regions. In the second part of my analysis, I find that effects of an increasing family day care share on child development are positive. Children seem to benefit in terms of their socio-economic maturity and motor skills. Including quality measures of family day care, however, matters. This indicates that reforms that rise the quantity of child care can have positive effects on child development if quality is regulated simultaneously.

This study provides evidence that important features of non-parental care at an early age are a small group size and a more family-like context. Due to that, family day

care might indeed represent a good option for harmonizing two public policy targets when flexibility is needed in expanding early child care in a fast manner: not to put too much of a burden on municipalities' financial situation and not to harm children's development.

However, one should keep in mind that the provision of this kind of care at a larger scale probably necessitates a different employment scheme for family day carers. At the moment, this care mode might not be attractive enough for staff to expand the family day care share by more in the future (Sell and Kukula, 2011). In a follow-up version of this study I will therefore look at the longer run impact of the child care reform in 2008 using most recent data and investigate the development of the family day care share until 2014. This future analysis will also include tests on the non-linearity of the effect of the family day care share.

Appendix

Table 4.8: Expansion of Center-Based Care over Time 2007-2011

Year	Child Care Centers per county	Child Care Centers Total (nationwide)
2007	110	45219
2009	113	46635
2011	116	47580

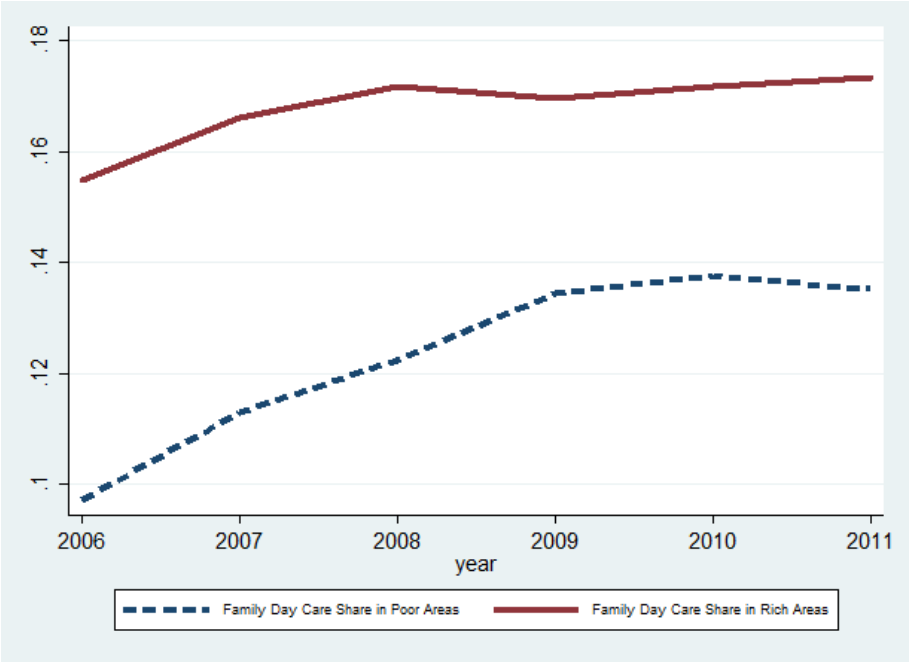
Notes: Center-based care is defined as care taking place in publicly subsidized child care centers that are monitored by district officials.
Source: Statistics of the *Child and Youth Services in Germany*. Own Calculations.

Table 4.9: Regional Characteristics over Time 2007-2011

Year	**2007**	**2009**	**2011**
Citizens per km2	519.808	518.350	519.650
	(672.276)	(672.963)	(678.946)
Living Space per inhabitant in m2	43.061	43.847	44.497
	(3.853)	(4.060)	(4.278)
Share 0-3-y-old children (in %)	2.432	2.421	2.402
	(0.255)	(0.245)	(0.256)
Log of GDPpc	3.270	3.263	3.201
	(0.332)	(0.314)	(0.304)
Unemployed (in %)	8.860	7.863	6.947
	(4.369)	(3.522)	(3.169)
Employed female (in %)	45.354	47.597	49.094
	(4.247)	(4.573)	(4.856)
Parttime-Employed female (in % of total working)	12.440	19.548	20.924
	(4.384)	(7.117)	(7.480)
To migration per 1000 inhabitants	38.447	39.474	40.318
	(14.822)	(15.269)	(16.054)
Out migration per 1000 inhabitants	38.972	40.304	38.375
	(11.750)	(12.893)	(12.440)
N	**411**	**411**	**411**

Notes: This table shows the regional characteristics (county averages) which are used as control variables in the main regressions displayed in Table 4.6. Standard deviation shown in parenthesis.
Source: INKAR. Own Calculations.

Figure 4.1: Family Day Care Share by County Income



Notes: This graph plots the family day care share over time for two groups of counties: The *poor group* (dotted) consists of all counties whose GDP per capita is below median GDP per capita in 2006. The *rich group* (solid) consists of the rest of counties.
Source: Statistics of the *Child and Youth Services in Germany* 2006-2011/INKAR. Own calculations.

Table 4.10: DiD Approach: Exploiting Different Expansion Paths due to Income Differences

	Family Day Care Share
After reform (post)	0.039***
	(0.013)
County GDPpc below median (treat)	-0.031***
	(0.007)
Interaction (post*treat)	0.019**
	(0.007)
Regional Controls	Yes
State FE	Yes
Year FE	Yes
Adj. R2	0.478
Observations	2299

Notes: Standard errors are clustered at the state-year level. *p < 0.10, ** p<0.05, ***p<0.010.
Source: Statistics of the *Child and Youth Services in Germany*/INKAR. Own Calculations.

Table 4.11: Regulations and Family Day Care Share: Full Results

Staff-per-child ratio in law	0.300**
	(0.114)
No Staff-per-child ratio rule[c]	0.038*
	(0.022)
Number of qualification hours[b] in law	0.004**
	(0.002)
No qualification hours rule[c]	0.067**
	(0.028)
Number of training hours in law	-0.004***
	(0.001)
No training hours rule[c]	-0.032
	(0.020)
Population density (citizens per km2 in 100s)	0.001**
	(0.001)
Living Space per inhabitant in m2	-0.003***
	(0.001)
Share 0-3-y-old children (in %)	0.014
	(0.013)
Share of 0-2y-olds in day care center or family day care(excl. 3y-olds)	-0.248***
	(0.047)
Share of church providers	-0.027***
	(0.010)
Share of other providers	0.002
	(0.012)
Log of GDPpc	-0.018*
	(0.010)
Unemployed (in %)	0.001
	(0.001)
Employed female (in %)	0.002***
	(0.001)
Parttime-Employed female (in % of total working)	-0.001***
	(0.000)
To migration per 1000 inhabitants	0.002***
	(0.001)
Out migration per 1000 inhabitants	-0.001
	(0.001)
Year: 2007 (D)	0.019*
	(0.011)
Year: 2008 (D)	0.037**
	(0.014)
Year: 2009 (D)	0.047***
	(0.014)
Year: 2010 (D)	0.055***
	(0.015)
Year: 2011 (D)	0.063***
	(0.016)
cons	0.249***
	(0.063)
State FE	Yes
Year FE	Yes
Adj. R2	0.477
Observations	2299

Notes: The dependent variable is the family day care share. [b]The number of required qualification hours is divided by 10 for the purpose of readability. [c]This binary variable equals 1 if there is no regulation on this quality indicator (otherwise it equals 0). Standard errors are clustered at the state-year level and are shown in parenthesis: * p< 0.10, ** p<0.05, *** p<0.01.

Source: Statistics of the *Child and Youth Services in Germany*/INKAR. Own Calculations.

Table 4.12: Effects of State Regulations on the Family Day Care Share – Alternative Specifications

	(1)	(2)	(3)
Staff-per-child ratio in law	0.300**	-0.476***	
	(0.114)	(0.161)	
No Staff-per-child ratio rule[c]	0.038*	-0.069**	
	(0.022)	(0.035)	
Number of qualification hours[b] in law	0.004**	-0.007***	-0.001
	(0.002)	(0.001)	(0.002)
No qualification hours rule[c]	0.067**	-0.040***	
	(0.028)	(0.011)	
Number of training hours in law	-0.004***	0.001	
	(0.001)	(0.001)	
No training hours rule[c]	0.025		
	(0.020)	(0.027)	
Regional Controls	Yes	Yes	Yes
State FE	Yes	No	No
Year FE	Yes	Yes	Yes
Adj. R2	0.477	0.402	0.212
Observations	2299	2299	260

Notes: The dependent variable is the family day care share. In the first column, results of the baseline empirical model (4.1) are shown. The dependent variable is the family day care share. In the second column, results of the the baseline empirical model (4.1) without the inclusion of state fixed effects are shown. In the third column, results of the the baseline empirical model (4.1) for the region of North Rhine-Westphalia are shown. In this state there was only a change of the stringency of the rule on required qualification hours. Therefore, relying only on variation across time (and not across states) in column (3), only the coefficient for this rule is estimated.
[b]The number of required qualification hours is divided by 10 for the purpose of readability.
[c]This binary variable equals 1 if there is no regulation on this quality indicator (otherwise it equals 0).
Standard errors are clustered at the state-year level and are shown in parenthesis:
* p< 0.10, ** p<0.05, *** p<0.01.
Source: Statistics of the *Child and Youth Services in Germany*/INKAR. Own Calculations.

Table 4.13: Which regional characteristics change simultaneously with the family day care share in North Rhine-Westphalia?

	Family-Day Care Share at Age 3
Citizens per km2 at age 3	0.0002
	(0.0003)
Employed female (in %) at age 3	0.004
	(0.0177)
Unemployed (in %) at age 3	-0.0073
	(0.0104)
Log of GDPpc at age 3	-0.1802
	(0.1557)
Parttime-Employed female (in % of total working) at age 3	-0.0077*
	(0.0042)
Living Space per inhabitant in m2 at age 3	-0.0848*
	(0.0435)
To migration per 1000 inhabitants at age 3	-0.1049***
	(0.033)
Out migration per 1000 inhabitants at age 3	0.0154
	(0.0349)
Share 0-3-y-old children (in %) at age 3	-0.007
	(0.0851)
Share of 0-3y-olds in day care center or family day care at age 3	-0.5196***
	(0.1819)
Pediatricians per 100 000 children at age 3	0.0026
	(0.0025)
% of employed persons holding university degree at age 3	0.0001
	(0.0003)
Self-employed per 100 employed persons at age 3	-0.0275
	(0.0509)
% of employed persons with immigrant background at age 3	0.0021
	(0.004)
Cohort 2011 (D)	0.0658*
	(0.036)
Cohort 2012 (D)	0.1373***
	(0.0457)
Cohort 2013 (D)	0.1994***
	(0.0547)
County Fixed Effects	Yes
N	409323
Adj. R2	0.8793

Notes: Standard errors are clustered at the county level and shown in parenthesis.
*p < 0.10, ** p<0.05, ***p<0.010.
Source: Statistics of the *Child and Youth Services in Germany*/INKAR. Own Calculations.

Table 4.14: Descriptive Statistics: Children at School Entrance Examinations in North Rhine-Westphalia (School Entrance Cohorts 2010-2013)

	Mean	**SD**
Child Development Indicators		
Adipositas (D)	0.046	(0.209)
School Readiness(D)	0.941	(0.236)
Socio-Emotional Maturity(D)	0.915	(0.279)
Language Skills(D)	0.685	(0.465)
Motor Skills(D)	0.829	(0.376)
Individual Characteristics		
Age (in months)	70.629	(3.533)
Male (D)	0.518	(0.5)
Birth weight (in gram)	2669.070	(1448.414)
Mom's education: basic	0.091	(0.288)
Mom's education: high	0.154	(0.361)
Nr of siblings (excl. kid)	0.895	(0.998)
Single parent	0.062	(0.241)
One/both parents foreigner	0.257	(0.437)
Supply of Early Child Care		
Share of Family Day Care at Age 3	0.189	(0.082)
Coverage All Provided Early Child Care at Age 3	0.134	(0.058)
Family Day Care Characteristics		
% of staff with completed family day care course at Age 3	0.668	(0.084)
% of family day care open more than 5 hours per day at Age 3	0.521	(0.133)
% of family day care with fewer than 4 children at Age 3	0.428	(0.134)
N	409323	

Notes: This table shows the descriptive statistics of children examined at the school entrance examination 2010-2013 in North Rhine-Westphalia (birth cohorts 2004/2005-2007/2008). The five child development indicators are measured as binary variables and equal 1 if the pediatrician does not detect a problem in the respective skill dimension (except in the case of adipositas: a value of 1 indicates a positive diagnosis of obesity).
Source: Statistics of the School Entrance Examinations in North Rhine-Westphalia and the *Child and Youth Services in Germany*. Own Calculations.

Table 4.15: Expansion of Early Child Care in North Rhine-Westphalia (2007-2011)

Year	Family Day Care Coverage	All Public Day Care Coverage	Share of Family Day Care	Share of Family Day Care (Minimum)	Share of Family Day Care (Maximum)	Family Day Carers per County (abs.)
2007	0.02	0.07	0.20	0.03	0.39	171
2008	0.02	0.11	0.16	0.01	0.43	201
2009	0.03	0.13	0.19	0.05	0.42	217
2010	0.04	0.14	0.21	0.07	0.40	231
2011	0.04	0.15	0.24	0.11	0.42	259

Notes: This table shows the average expansion of early child care in North Rhine-Westphalia from 2007-2011. Minimum and Maximum values display the smallest resp. the largest county average of the family day care share.
Source: Statistics of the *Child and Youth Services in Germany*. Own Calculations.

Chapter 5

Class Composition and Educational Outcomes – Evidence from the Abolition of Denominational Schools

This chapter is joint work with Ilka Gerhardts from the University of Munich (LMU) and Uwe Sunde from the University of Munich (LMU).

5.1 Introduction

Many schools are operated on a basis of multigrade classes, which represent the typical way of teaching children in the context of limited resources. Such multigrade classes are a cost-effective way of providing children with education. In fact, in large parts of the world schools with multigrade classes, often run by different religious denominations, represent the typical way of teaching children. Around the globe, approximately one third of all classes across all countries, including some of the more developed countries, are multigrade classes (2005 UNESCO Agenda for Educational Planning).

In the face of the dramatic demographic change, multigrade classes have recently become a principal adjustment device for enrollment fluctuations also in many parts of Europe. However, warnings have been raised regarding the potentially detrimental effects of teaching students of different ages and maturity within the same room. At the same time, denominational affiliation has lost importance, and demographic change led to the abolition of denominational schools. Mixed empirical evidence regarding the effects of abolishing denominational schools with multigrade classes on subsequent outcomes fuels heated debates regarding the appropriate education policies.

This paper investigates the impact of denominational schools on the returns to education. The identification strategy exploits the natural experiment of the abolition of denominational schools in the Saarland, a state in Germany, in 1969. Prior to the reform, more than 95% of primary and lower secondary schools were church-maintained. In scarcely populated regions, the strict tracking by religious denomination imposed severe restrictions on the allocation of students. As a consequence, schools were relatively small, implying that students of different ages and skills were taught within the same classroom, i.e. in multigrade classes. The abolition of denominational schools in 1969 led to the dissolution of hundreds of these rural multigrade schools within less than a year. The remaining schools obtained a single-grade structure, similar to the larger schools in more urban environments.

Our identification approach exploits differential treatment exposure of students depending on how many students of the same birth cohort have the same denomination. In more rural municipalities multigrade teaching in denominational schools was the norm prior to 1969, but not afterwards. By contrast, in more urban municipalities multigrade teaching in denominational schools was not necessary due to higher student numbers. To estimate the effects of the reform on schooling and labor market outcomes we use an enhanced differences-in-differences approach.

By exploring the heterogeneity of the effects across gender, the evidence also provides new insights into the roots of gender inequality. In particular, the large-scale natural experiment enables insights into the socialization mechanisms at school that might

lead to gender differences in labor market participation and occupational choice later on in life.

The empirical analysis is based on a unique combination of administrative records and comprehensive population census data. The dataset has been collected and digitized specifically for this research project, which to our knowledge is the first to exploit the abolition of denominational schools as a natural experiment in this context. Using municipality codes and schools' denominations, we are able to link individual-level census data on virtually all of Saarland's households in 1970 and 1987 to a comprehensive schools' index that comprises more than 7,500 school-year observations on a municipality-denomination-level. The availability of a wide range of schooling covariates allows us to control for channels like class size, school size, school consolidation, gender composition, etc. that might confound the multigrade effects.

The empirical results suggest that multigrade classes have detrimental effects on final grade attainment and labor market participation. While all students profited from the abolition of denominational schools in terms of the higher grade attainment and a greater likelihood to become a white-collar worker, the effect is notably stronger for girls. The abolition of denominational schools in municipalities where multigrade teaching was the norm before 1969 led to an increase in the number of girls who attained a higher educational degree and a decrease in the number of girls becoming housewives. The results therefore suggest an interplay of gender socialization and the mode of teaching in terms of multigrade classes on subsequent outcomes.

The question how denominational schools with multigrade classes affect students' outcomes touches upon several research strands related to class composition, educational infrastructure, peer and tracking studies. Our empirical approach contributes to the literature in several ways. First, the natural experiment of the sudden abolition of denominational schools allows for a credible identification of the causal impact of denominational schools with multigrade classes, whereas many existing studies suffer from insufficient randomization which renders identification problematic (mainly because of self-selection). Second, we present effects that are placed in a

Western European society. By contrast, those studies on multigrade classes with credible identification (due to controlled randomization) have been conducted mainly in developing countries, at the cost of limited external validity for more developed countries. Third, the high-quality dataset covering virtually the complete population of our region of study minimizes selection and response biases and affords statistical power whereas existing research mostly relies on evidence from small samples. Fourth, provided with large-scale evidence, we are able to link gender mechanisms at school not only to final grade attainment but also to labor market participation and occupational choice. Our analysis thereby extends earlier work that mainly focused on the gender specific effect of class composition on schooling outcomes. Overall, our results are in line with the findings of earlier studies that suggest rather negative effects of multigrade classes.

The remaining part of the paper is structured as follows. Section 5.2 gives an overview of the existing literature on class composition. Section 5.3 describes the institutional background. Section 5.4 presents the identification strategy, followed by a compact presentation of the data in Section 5.5. Section 5.6 presents the empirical results, discusses robustness with respect to sensitivity checks and shows the results of the subgroup analysis. Section 5.7 concludes.

5.2 Literature Review

Multigrade classes[1] produce multiple forms of *peer effects*. Peer effects are central aspects of education research. They have been modeled as inputs to the education production function ever since Coleman (1968) made them popular, among others by Iversen and Bonesrønning (2015) and Jones (2013). There exists relatively less research on peer effects of class composition than, e.g., on class size (Jones, 2013), but the absolute number of class composition studies is still vast. Many of those have been criticized for low methodological quality, however, as detailed in Johansson and

[1] Multigrade classes, as opposed to single-grade classes (Veenman, 1995), do not sort students by age and skill. Furthermore, they are created out of some necessity, not pedgogical purpose, as other types of combination classes are.

Lindahl (2008) or Mason and Burns (1996). In general, a variety of peer effects can arise in a system of multigrade classrooms which has been touched upon as follows.

Between-student spillovers may be positive if more knowledgeable, skilled or able classmates serve as natural role models (Duflo *et al.*, 2011; Hanushek *et al.*, 2003). Practical relevance of peer collaboration, however, is told to be rather limited (Hattie, 2002). There is also evidence that peer effects are rendered negative if age gaps arise due to grade repeating and redshirting which is often the case in developing countries (Lavy *et al.*, 2012; Jones, 2013).

Finally, peer effects among teachers in the sense of shared experiences have been mentioned in the multigrade context. The probability of beneficial spillovers prerequisites at least two teachers per school and is likely to increase in larger teaching staff which puts rural schools at a disadvantage (McEwan, 2008).

Besides peer effects, also effects of (no) adjustments of teacher training, curricula, materials and incentives need to be reconsidered upon collapsing the grade level structure. Traditional teacher colleges prepare single-grade teaching although multigrade teaching is strategically more demanding and stressful (Mason and Burns, 1996; Russell *et al.*, 1998). Therefore, it is likely that multigrade schools have negative effects on students if the pedagogical infrastructure is not adapted to multigrade teaching.

Current research on multigrade classes is frequently located in developing countries. See Little (2001) or McEwan (2008) for overviews in Africa, Asia and Latin America respectively. While some randomized control studies conducted in these countries convince by providing internal validity, their external validity is rarely given.[2] First, there are several institutional deficiencies that make it difficult to compare the examined multigrade settings to each other. For example, in some cases the mixed grade levels are not even adjacent (Mulkeen and Higgings, 2009) which increases the hetero-

[2] Not only randomized control studies deliver evidence for multigrade effects in developing countries. Jones (2013) relies on an IV strategy to circumvent selection issues. He presents strongly negative effects by African overage-for-grade peers thus being supportive of Lavy *et al.* (2012).

geneity in the classroom substantially.[3] Second, unsafe school ways complicate school attendance asymmetrically for girls which changes the classroom gender distribution (Mulkeen and Higgings, 2009). Third, grade attainment may not mean anything regarding knowledge and skills (Jones, 2013). Due to this range of peculiarities in developing countries estimation of the effects of multigrade classrooms is challenging even when applying (quasi-)experimental designs that are good practice in the sense of Angrist (2004).[4]

Even though the major part of research on multigrade classes studies multigrade settings in development countries multigrade classrooms are also prevalent in more developed countries. Contemporaneously, multigrade classes make up one third of all classes on earth, and even in countries like Finland, the Netherlands, India, Peru, Sri Lanka and Pakistan multigrade predominate single-grade classes (Mulkeen and Higgings, 2009).

Existing studies on multigrade classes that were (mostly) conducted in industrialized countries up to 1995 are summarized in a meta-analysis by Veenman (1995). He concludes that there are no significant effects on cognitive and/or social-emotional outcomes after averaging over 43 combination class studies meeting his econometric criteria. Apart from being quite outdated today these criteria were already criticized by contemporary scholars Mason and Burns (1996). They point out that Veenman (1995) draws on studies that use non-random samples. They argue that multigrade classes have better teachers and students. By that the group composition in multi-

[3] Furthermore, teachers in these countries often undergo very different trainings and the rate of teacher absence is very high. Enrollment is not compulsory but rather an achievement in itself, at any age (Jones, 2013).

[4] Vivalt (2015) establishes the overall limited external validity of impact evaluation studies formally.

grade classrooms biases an actually negative effect of less effective teaching in this setting towards zero.[5]

A rather recent study on combination classes is the one by Johansson and Lindahl (2008). They rely on survey data and compare non-random but observationally equivalent single-grade and mixed-age classes in Sweden. They report a negative impact as sizable as that observed for larger classes in the STAR experiment.[6] Another recent approach to estimate effects of multigrade classrooms is presented by Leuven and Rønning (2014). Looking at multigrade schools in Norway they highlight the idea of *perspective-dependent* peer instruments obtaining contrastive signs out of the same data. They find younger students to benefit from having older ones around while older students get worse results when younger ones are around.[7] Leuven and Rønning (2014) conclude seemingly inconsistent evidence to be rooted in researchers' unilateral approaches. Furthermore, they claim to reconcile the literature finding small but significantly positive peer effects conditional on an optimal allocation.[8] Subsequent investigations by Carrell *et al.* (2013), however, point out limitations of peer group interventions as proposed by Leuven and Rønning (2014) in the face of endogenous subgroup formation. They deliberately allocate weak and strong ability students enabling theoretically the largest possible spillovers. They do not foresee

[5] Concretely, multigrade teaching is found to cover less curriculum, especially in higher grades. Russell *et al.* (1998) back up the hypothesis that multigrade teaching is increasingly detrimental beyond basic skill acquirement. Furthermore he finds numeracy skills to suffer more than literacy skills from a multigrade structure in elementary schooling. To the extent of bias due to peer ability Mason and Burns (1996)'s critic is mitigated by Cullen *et al.* (2006). They present evidence from US school choice lotteries claiming no significant influence on student attainment by higher peer quality associated with the preferred schools. Their quality indicator measures the difference between (single-grade) classmates' average test scores after winning or loosing the lottery. Insignificance applies uniformly to ability, gender and race strata. It is also robust to all intensities of lottery-induced peer improvement.

[6] In the STAR framework the presence of about six more students reduces test scores of classmates by 4 percentage points in the first year and 1 additional percentage point in subsequent years (Krueger, 1999).

[7] Concretely, they refer to Jacob *et al.* (2010) deriving negative impacts from measuring exposure to lower grade levels thus taking the perspective of the harmed older students. Along the same pattern Thomas (2012) is expected to find positive peer effects because he considers higher grade levels that are taught together with the treated younger students.

[8] Similarly Duflo *et al.* (2011) uncover opposing spillover effects for high and low achievers in Indonesian (single-grade) schools. However, after taking into account lasting consequences of more adequate curricula (detailed in Glewwe *et al.* (2009)) and teachers' tendency to teach to the top of the class, Duflo *et al.* (2011) find tracking to be beneficial for all students. Yet another (single-grade) example where curriculum adjustments persistently outweigh peer effects is presented by Cortes and Goodman (2014) looking at US schools.

more able students to cut less able ones out of their circle leaving them with even worse academic attainments.

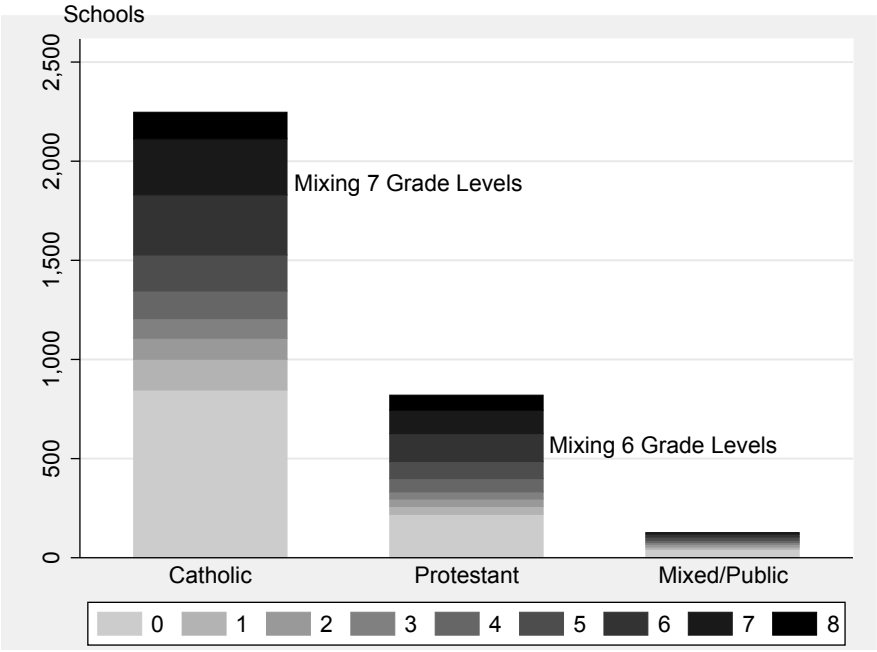
In view of the existing research on multigrade classes our study contributes to the literature in several ways: Our study focuses on the impact of the multigrade setting in German schools and uses a natural experiment – the sudden abolition of denominational schools – for identification of the causal effect of multigrade schools. By contrast, existing studies like those of Johansson and Lindahl (2008) and Leuven and Rønning (2014) suffer from insufficient randomization and rely on selection-on-observables methods which render causal identification problematic. Furthermore, we present effects of multigrade classes that are placed in a Western European society while those studies on multigrade classes with credible identification have been conducted mainly in developing countries. But, as described above, there are quite a few limitations of the institutional settings in these countries which diminish the external validity of the findings for industrialized countries. Additionally, we possess a high-quality dataset covering virtually the complete population of our region of study. Thus, we do not have to deal with selection and response biases as much as studies relying on survey data (such as Johansson and Lindahl (2008)). Another advantage of being provided with large-scale evidence is that we are able to explore the effects of multigrade classrooms not only with respect to final grade attainment (as most existing research is confined to) but also to labor market participation and occupational choice. Extending the multigrade analysis to an interplay of medium-run outcomes (as pioneered in other contexts by Clark and Del Bono (2016) and Greenwood *et al.* (2016)) is new to the literature.

5.3 Institutional Background

This section describes the school reform in the region of our study, the framework of schooling laws, as well as potential confounders, using information from various sources.

Prior to the reform in 1969, almost all *Volksschulen* sorted students by denomination. This allocative restriction created multigrade classes in regions with a low population density. Figure 5.1 provides a first overview of the prevalence of multigrade classes in the Saarland prior to the reform.[9] With few exceptions denominational schools played a role only in the lowest educational track. For a concise description of ability tracking in German schools see Pischke and Wachter (2005).[10]

Figure 5.1: Mixed Grade Levels by Denomination



Notes: This figure shows the prevalence of multigrade teaching prior to the reform in 1969 by denomination. Each shade represents the amount of grade levels that were taught together. Black, for instance, shows the number of schools that were teaching 8 grade levels simultaneously.

Source: Schools' Index 1964-1986. Own calculations.

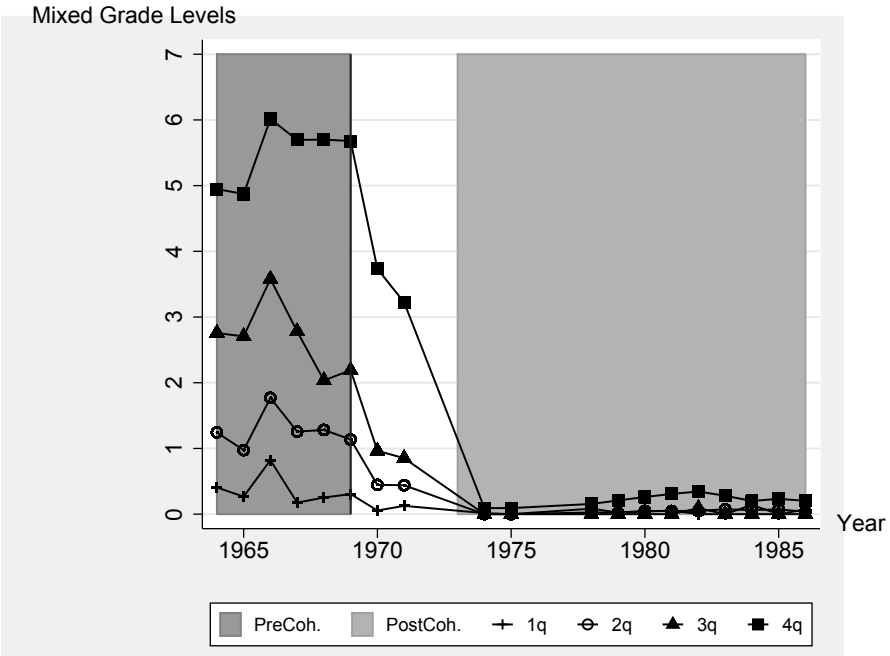
[9] Rural *Volksschulen* create a multigrade setting not supported by pedagogical adjustments. First, the schools' records do not provide any evidence for adjustments. Moreover, albeit this is no rocket-science, there do exist alarming hints about amateurishly adapted teaching practices, available at http://www.spiegel.de/spiegel/print/d-46265072.html (01 May 2015) which highlights the comparability problem to mixed-age classes (Mulkeen and Higgings, 2009).

[10] Multigrade classes in remote regions pool children of very different abilities. Do the observed spillovers of our study provide guidance for inclusion of handicapped children as well? This depends on the multigrade school employing a full inclusion policy. Iversen and Bonesrønning (2015) explore spillovers in Norwegian elementary schools where special education happens to be integrated within ordinary classrooms. They find that spillovers interact with the level of special education provided. In Germany the *Volksschule* and special schools are kept apart. After reforming lower secondary education the separation persists (Figure 5.4). Thus the insights by Iversen and Bonesrønning (2015) formalize the lack-of-comparability argument forwarded in Veenman (1995) by which he excludes studies on gifted as well as handicapped children from his synthesis.

Schools providing primary or lower secondary education were uniformly labeled *Volksschule*, see Figure 5.4 in the appendix for a more details on the distribution of school types over time.

Prior to the abolition of denominational schools, the treatment exposure (the probability of being taught in a multigrade school) of students was dependent on how many students of the same birth cohort had the same denomination – due to the legal obligation to teach Catholics and Protestants separately.[11] In sum, 75% of schools in the Saarland resolved to a multigrade structure prior to the reform in 1969, all of which were schools in more rural regions. Denominational schools in more urban regions, by contrast, were characterized by a single-grade structure.

Figure 5.2: Mixed Grade Levels by Treatment Probability over Time: Catholic Students

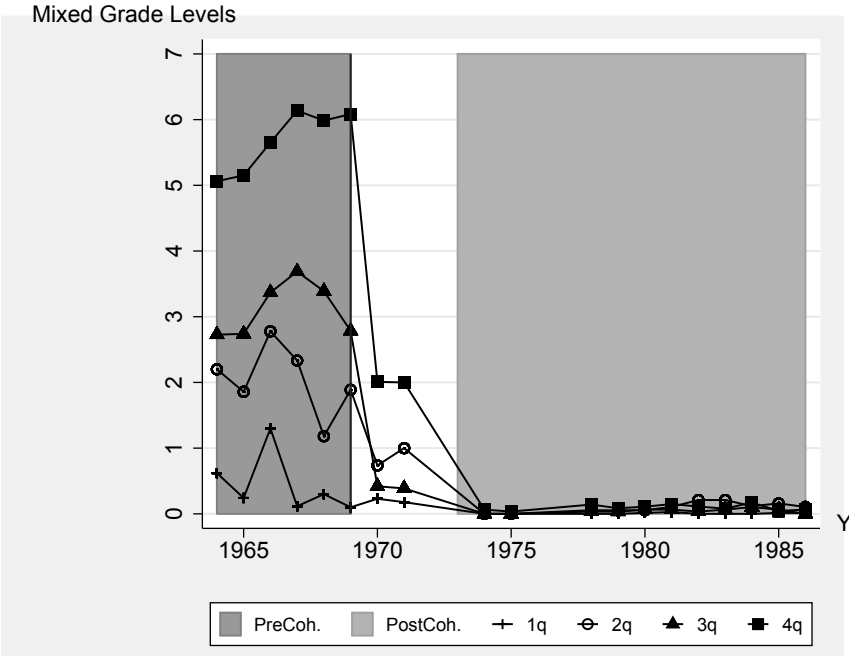


Notes: This figure shows, for the case of Catholic students, the prevalence of multigrade teaching (diyplaying the number of mixed grade levels) over time by treatment probability (in quartiles). The treatment probability depends on the number of schools in a municipality-denomination-cohort-cell that were offering multigrade teaching prior to the reform.

Source: Schools' Index 1964-1986. Own calculations.

[11] Verfassung des Saarlandes (1947) Art. 27 (Amtsbl. des Saarlandes, Nr. 41) Vom 05.11.1969, *available at* http://www.verfassungen.de/de/saar/saarland47-index.htm (23 May 2015).

Figure 5.3: Mixed Grade Levels by Treatment Probability over Time: Protestant Students



Notes: This figure shows, for the case of Protestant students, the prevalence of multigrade teaching (diyplaying the number of mixed grade levels) over time by treatment probability (in quartiles). The treatment probability depends on the number of schools in a municipality-denomination-cohort-cell that were offering multigrade teaching prior to the reform.

Source: Schools' Index 1964-1986. Own calculations.

The reform of 1969 had a direct impact on schools offering basic education. Inducing a change in students' distribution across school types it also indirectly affected higher education though. When denominational schools were legally abolished in various states all over Germany, this raised hot debates and interventions on behalf of the church and parents likewise[12], but in the Saarland the reform was carried out neatly.

Due to the reform the number of multigrade schools decreased by two thirds in less than a year and from 1974 onwards the share of multigrade schools was negligible. Thus, the reform changed the learning environment for children in more rural regions where multigrade schools predominated prior to the reform in 1969 substantially. Tiny schools were wrapped up into normal-size ones reducing the number of village schools by more than 50% while diminishing the frequency of more urban schools

[12] http://www.spiegel.de/spiegel/print/d-46369565.html (01 May 2015).

only moderately. In consequence, from 1974 onwards the prevalence of multigrade teaching was close to zero in both treated and control regions, see Figure 5.2 for the development of multigrade teaching in Catholic schools over time and Figure 5.3 for the case of Protestant schools.[13]

The abolition of denominational schools left some villages without an own school altogether and required their children to become commuters. Having to commute anyway changed relative commuting costs to higher education schools that might previously have been prohibitive. Attending a restructured *Volksschule* or even opting for a higher education school, either way rural students were taught in much more homogeneous classes.

Table 5.1: School Characteristics by Treated and Control Status of Catholic Students

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Class Size	37.509	34.42	-3.089	(-9.987)	23.24	21.701	-1.539	(-8.478)
Students/Teacher	36.354	34.621	-1.733	(-5.678)	20.076	21.002	.926	(4.471)
Students/School	369.435	109.818	-259.617	(-47.896)	284.636	127.83	-156.806	(-28.302)
Girls' Share	.527	.49	-.037	(-6.04)	.48	.49	.01	(4.738)
Female Teachers' Share	.459	.427	-.033	(-3.965)	.526	.529	.004	(.471)
Teachers/School	10.125	3.056	-7.069	(-48.393)	14.292	6.155	-8.137	(-29.917)
Observations	1216	1021			2667	872		

Notes: A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. In this table, only Catholic students and the schools they attended are considered.
Source: Schools' Index 1964-1986. Own calculations.

All key features of schools are summarized in Tables 5.1 and 5.2, partitioning the universe of *Volksschulen* into four groups, namely treated and control schools, each before and after 1969 (and separately for Catholic and Protestant schools).[14] As the tables show, by construction the reform reshaped the educational infrastructure in multiple ways and also implied more students and more teachers per school in absolute terms (EENEE, 2015). For example in the case of Protestants living in

[13] Tables 5.6, 5.7 and 5.8 in the appendix compare the number of mixed grade levels in treated and control regions prior and after 1969 separately for Catholic, Protestant and (the few) non-denominational schools.

[14] The key features of non-denominational schools are shown in Table 5.9 in the appendix.

Table 5.2: School Characteristics by Treated and Control Status of Protestant Students

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Class Size	32.409	31.009	-1.4	(-3.193)	23.045	22.344	-.7	(-3.913)
Students/Teacher	31.465	31.404	-.061	(-.105)	20.336	20.215	-.122	(-.599)
Students/School	270.118	88.962	-181.156	(-27.511)	252.835	226.88	-25.955	(-4.335)
Girls' Share	.51	.494	-.016	(-1.933)	.483	.483	0	(-.06)
Female Teachers' Share	.517	.463	-.054	(-4.265)	.526	.528	.002	(.288)
Teachers/School	8.607	2.772	-5.835	(-27.904)	12.599	11.414	-1.185	(-3.986)
Observations	374	448			2607	932		

Notes: A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. In this table, only Protestant students and the schools they attended are considered.
Source: Schools' Index 1964-1986. Own calculations.

treated municipalities where multigrade teaching was the norm prior to the reform, average school size increased from 89 students per school to 227 students per school and from 2.8 teachers per school to 11.4 teachers per school (see Table 5.2). At first sight surprisingly, average class size shrank because the inflow of remote area children into more urban school districts was mitigated by a demographic decline in enrollment. It drastically reduced overall class size from 39 (1964) to 19 (1986) students on average, but the relative change was identical for treated and control regions. However commuting students coming from remote areas might have encountered higher quality peers from more urban municipalities (Leuven and Rønning, 2014).

For the comparison between treated municipalities (where multigrade teaching was the norm prior to 1969) and control municipalities (where single-grade teaching was the norm prior to 1969) to make sense a common trend between those regions is essential. The 1960s are called the decade of educational expansion and changes over time are indeed tremendous. We exploit that the reform eradicates multigrade classes which creates an asymmetry between otherwise parallel worlds. The following important education laws in the Saarland are all implemented well before the reform is rolled out in 1969 and they maintain a common denominator for treated and control municipalities – those with and those without a history of multigrade schools – over time.

To begin with the *Compulsory School Entry Age* fixes enrollment into primary school to age six with minor exceptions referring to each June's 30th as cut-off date.[15] Next *Compulsory Schooling Duration* requires that students stay in school for at least nine years and passing the ninth grade is rewarded with a lower secondary degree. It turns out that roughly 4:1 students finish a ninth grade already before the law inures in 1965 (Pischke and Wachter, 2005). Its implementation, however, requires two short school years that actually compress schooling duration in 1966/67. Then, *No Tuition Fees* guarantee basic education to be free of charge, independent of the school being state- or church-maintained.[16] It limits the influence of parents' financial constraints and prevents a selection by the fee itself. Finally, *Limited School Choice* of the parents is achieved by allocating students over schools based on catchment areas.[17] To choose a certain *Volksschule* by its reputation would require the household to move into that school's catchment area. Rothstein (2006) investigates parental preferences in school choice and establishes that peer groups matter even more than schools' effectiveness. This underlines the importance of student allocation by catchment areas because it mitigates parental choice effects which interfere with the core mechanism of multi-grade classes. Jointly these laws provide accuracy in comparing schooling circumstances. This is an advantage compared to class composition studies of developing countries.

We analyze a period of more than two decades of schooling conditions. Our setup is robust to symmetric shocks. When screening the most influential historical events that could have had asymmetric impacts on treated and control municipalities, a primary concern relates to fluctuations in economic activity centered in urban regions. The coal and steel crises depressed the Saarland even more than the rest of Germany (Lichtblau, 2009). They caused dramatic peaks in unemployment and overshadowed positive shocks such as the construction of the Ford plant or the infrastructure improvement by the Saar Canal. Geographic controls measuring the distance to former major smelting works, direct access to the river, etc. are one possible solution to

[15] §2 Satz 1 Gesetz Nr. 826 Schulpflichtgesetz *available at* http://sl.juris.de/cgi-bin/landesrecht.py?d=http://sl.juris.de/sl/gesamt/SchulPflG$_S$*L.htm#SchulPflG*$_S$*L*$_r$*ahmen*(12*June*2015).

[16] §1 Satz 1 Gesetz Nr. 662 Schulgeldfreiheit *available at* http://sl.juris.de/cgi-bin/landesrecht.py?d=http://sl.juris.de/sl/gesamt/SchulGFrhG$_S$*L.htm*(12*June*2015).

[17] §29 Satz 2 Schulordnungsgesetz vom 5. Mai 1965.

control for these changes. It is worth mentioning that despite of these shocks the Saarland was politically nearly perfectly stable (ibid). Only the very last year of our study's time horizon is subject to a different government, therefore we expect its influence to be limited. The advantage of exploring inner-state differences becomes obvious here. By construction, many complicating aspects like tax schedules causing potential problems in Abramitzky and Lavy (2011), etc. are taken care of from the start.

5.4 Empirical Model

The key empirical question refers to the comparison of the performance of students in a multigrade environment to a single-grade environment, which is less heterogeneous in terms of birth cohorts. We tackle this question estimating a triple differences (DDD) model that exploits exogenous variation in the probability to be a multigrade student over time, region and age group.
Let Y_{1imdcy} represent individual i's outcome in municipality m with denomination d, belonging to cohort c and age group y if she attended a multigrade school and Y_{0imdcy} otherwise.

A student is defined as treated if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. If in one municipality there was one Protestant school teaching at least two grade levels jointly in all pre-reform years, then a Protestant student will be labeled as living in a multigrade municipality. This is still true if in the same municipality there exist Catholic schools which might be single-grade schools. This definition underlies the balancing tables 5.1, 5.2 and 5.4. It ensures that within a treatment-municipality-denomination-cohort cell the probability to attend a multigrade school was 100%.[18]

[18] We estimate an intentention-to-treat effect. Apart from the standard assumptions for multiple differences analyis our setup requires two non-technical assumptions. First, pre-reform denomination of student and school coincide and second, the likelihood for treated and control students to start their own household follows a common trend while they are under-age. Conditional on these assumptions the probability to be treated assigned by the binary multigrade indicator is 100%.

Yet, this definition might be overly retrictive as it dismisses multigrade exposure whenever the probability was not 100%. In other words, citing the example from above, even if only in one year prior to the reform the Protestant school obtained a single-grade structure the Protestant student will be labeled as non-treated. Therefore, building on the binary defnition we employ two alternative continuous treatment indicators in our regressions.[19] Consider a municipality with two Protestant schools, school A with 90 and school B with 10 students. The school-based indicator corresponds to the share of multigrade schools, the student-based indicator to the share of multigrade students of the respective municipality-denomination-cohort cell. Table 5.3 shows which indicator behaves more conservative, in the common computational scenarios.

Table 5.3: Treatment Status by Alternative Multigrade Indicators

	Multigrade Indicator		
Multigrade School?	Binary	School-based	student-based
Case I			
Both A, B	1	1	1
Case II			
School A	0	0.5	0.9
Case III			
School B	0	0.5	0.1
Case IV			
Neither A nor B	0	0	0

Note: Fictitious example considering a municipality with two Protestant schools, school A with 90 and school B with 10 students. The continuous school-based indicator corresponds to the share of multigrade schools, the continuous student-based indicator to the share of multigrade students of the respective municipality-denomination-cohort cell.

The binary indicator underlying our balancing tests is very conservative in assigning treatment status. Thus, it is most likely to reveal significant differences that potentially create non-common trends. Nevertheless, as any binary indicator, it disregards that treatment probability is gradual. Therefore it should be modeled as a continuous variable, just as we do in our preferred specifications discussed in this paper. As Table

[19] The binary treatment indicator is used in a robustness check. The results do not provide additional insights and are available upon request.

5.3 shows the school-based indicator computes the probability to attend a multigrade school based on the number of schools per municipality-denomination-cohort cell (MDC). The student-based indicator models the probability to attend a school within a MDC cell to be proportional to the school's size, as a proxy for its capacity to take in students. Note however that the latter need not be a better indicator per se. Smaller multigrade schools were often much more extreme in collapsing grade levels than larger schools had to be. This motivates to condition on treatment intensity, something we are still working on. Of course treatment probability and treatment intensity are two different things. This is just one example to point out that apart from school size there exist multiple factors influencing the possible multigrade experience of a student. From this perspective, the school-based indicator is just a neutral and thus very useful benchmark.

We estimate the *reform effect* in a regression with $Multigrade_{md} \in [0,1]$, a continuous variable measuring the likelihood of being taught in a multigrade class, the binary variable $c \in \{Pre, Post(Reform)\}$ and the binary variable $y \in \{Young, Old\}$, and a triple interaction, reflecting the DDD estimator. *Post* equals one for observations of the 1987 Census and zero for 1970. *Young* equals one for people aged fifteen to twenty in either census year and is zero for people aged 32 to 37 years.

$$\begin{aligned} Y_{imdcy} = \beta_0 &+ \beta_1 Multigrade_{md} + \beta_2 Post_c + \beta_3 Young_y \\ &+ \beta_{12} Multigrade_{md} Post_c + \beta_{13} Multigrade_{md} Young_y + \beta_{23} Post_c Young_r \\ &+ \beta \underbrace{Multigrade_{md} Post_c Young_y}_{D_{mdcy}} + \psi_m + \epsilon_{imdcy} \end{aligned} \quad (5.1)$$

To account for time-invariant confounders at the municipality level, we include municipality fixed effects ψ_m. To allow for correlation of errors within municipality we cluster on the municipality level (335 clusters).

Identification is thus based on the contrasts across time, age groups and municipalities with a different coverage of multigrade schools prior to the reform. We estimate the DDD baseline reform effect including just the main effects *Multigrade*, *Post*, *Young* and their interaction terms.

We proceed by estimating the multigrade effect in more extensive specifications that include additional individual controls from population census data. These include *Age*, *Age Square*, *Young at School Entry*, *Female*, *Catholic* and *German*. *Young at School Entry* relates birth month and school entry cutoff date to indicate if a student is relatively young within her cohort. Combining this with administrative data from school records allows us to include additional controls. These comprise municipality-denomination-cohort level regressors *Class Size*, *School Size* (defined as the number of students) *Girls' Share* and *Female Teachers' Share*. We furthermore account for *Potential Commuting Costs* which we define as the average distance to the nearest *Realschule* or *Gymnasium* net of the distance to the nearest *Volksschule*.

The identifying assumption of our DDD strategy is that multigrade exposure is as good as randomly assigned conditional on observables and unobservable-but-fixed confounders. Adding a control group of elder people nets out region-specific changes that are not rooted in schooling conditions themselves. An example would be a boost in multigrade municipalities' neighborhood quality induced by state-level interventions to counteract drift to the cities (characterized by single-grade schools). The setup still requires unobservable asymmetries in teaching effectiveness and ability differences between multigrade municipalities' and single-grade municipalities' students to be time-constant, because – with only two periods in which region-specific outcomes are measured – trends are not identified, a drawback detailed in Stephens and Yang (2014). Moreover we rely on the aforementioned student allocation via catchment areas to ensure that students do not choose their school, and thus their multigrade exposure. To sum up, for multidimensional differencing to be applicable group composition needs to be spatially stable as well as groups should follow a common trend over time. Furthermore we assume *zero conditional mean*, *additive separability* and a *constant, weakly monotone causal effect* β.

5.5 Data

This section describes the data. Via municipality codes we combine two censuses and one schools' statistics, all of which are comprehensive, high-quality administrative datasets.[20]

Outcomes[21]

We construct schooling and labor market outcomes using individual-level census data from 1970 for the baseline and from 1987 for the follow-up cohorts. The data is available via remote execution at the German Federal Statistical Office. To evaluate final grade attainment we consider two separate dummies, namely (1) attainment of *Mittlere Reife or Fach-/ Abitur* (i.e. at least an intermediate secondary degree) and (2) attainment of *Fach-/ Abitur* (i.e. at least a high-school degree). Looking at grade attainment instead of years of schooling reflects longer schooling net of grade repetition and also identifies dropouts (EENEE, 2015). There are no test scores in the data. If there were, however their predictive power might have been limited anyway by grading on a reference curve, especially in a multigrade class, because relative grading depends on the presence of more advanced peers (Leuven and Rønning, 2014). Importantly, peer effects may trigger social competences not captured by test scores but perhaps reflected in post-schooling attainment. We therefore also use labor market outcomes to assess lasting or reemerging effects of schooling similar to Chetty *et al.* (2014b). In order to analyze labor market participation we use binary indicators on unemployment and labor market participation. Given labor market entry we distinguish further between blue- and white-collar occupations to capture the socio-economic status of the occupation. Note that wages are not reported in the Census 1987.[22] Table 5.4 shows descriptive evidence on the differences between treated and

[20] Volkszaehlungsgesetz 1970 vom 14. April 1969 (BGBl. I S. 292); Volkszaehlungsgesetz 1987 vom 8. November 1985 (BGBl. I S. 2078).

[21] Nearly all our outcomes are binary. Accordingly, the OLS regressions represent linear probability models (LPMs) which means that causality draws on the CIA, predictions may violate the [0,1] range and the error term is heteroskedastic (Angrist and Pischke, 2008).

[22] For a follow-up version of this paper, we consider to assign a standard income range based on each observation's meticulously reported profession (ISCO 88) for income mobility analysis in the sense of Chetty *et al.* (2014a).

control individuals with respect to their schooling and labour market outcomes. It shows that treated individuals prior to the reform were less likely to hold at least a *Realschule* degree (RS degree) than control individuals. Furthermore, they were more likely to have a blue-collar job and less likely to have a white-collar job. According to the descriptive statistics, these differences were less pronounced after the reform. In fact, after the reform treated individuals are more likely to hold at least a *Realschule* degree than control individuals.

Table 5.4: Descriptive Statistics: Treatment, Outcomes and Controls

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Treatment Indicators								
MDC MG School Share	.259	1	.741	(397.086)	.028	.122	.094	(59.844)
MDC MG Pupil Share	.088	1	.912	(821.797)	.005	.064	.06	(53.115)
Outcomes								
At least RS Degree	.094	.08	-.014	(-5.298)	.371	.392	.021	(3.751)
At least A-levels	.009	.007	-.002	(-1.834)	.067	.069	.002	(.614)
Employed	.651	.653	.001	(.328)	.688	.707	.019	(3.674)
Non-Participant LM	.071	.07	-.001	(-.349)	.045	.032	-.013	(-5.694)
Blue-Collar Job	.514	.548	.034	(7.485)	.525	.538	.013	(2.313)
White-Collar Job	.407	.364	-.043	(-9.674)	.428	.428	0	(-.019)
Controls								
15-17 Year-olds	.417	.43	.013	(2.919)	.218	.227	.009	(1.966)
1 VS in MDC cell	.297	.902	.604	(156.353)	.325	.82	.495	(97.561)
Mun: max.5000 inh.	.233	.882	.649	(178.032)	.307	.893	.586	(121.127)
Female	.498	.488	-.011	(-2.376)	.449	.435	-.014	(-2.442)
Age	17.846	17.794	-.052	(-3.58)	18.566	18.518	-.048	(-3.276)
Young Within Cohort	.396	.402	.007	(1.489)	.372	.38	.008	(1.513)
Catholic	.804	.692	-.112	(-30.104)	.804	.682	-.123	(-26.083)
Protestant	.187	.292	.106	(28.829)	.17	.277	.107	(23.929)
German	.967	.979	.012	(7.898)	.952	.968	.016	(7.055)
Single	.895	.893	-.002	(-.75)	.944	.951	.007	(2.654)
Household Size	4.376	4.65	.274	(15.244)	3.742	4.039	.297	(19.378)
MDC Class Size	37.037	34.447	-2.59	(-77.175)	23.215	22.337	-.878	(-52.595)
MDC Pupils	380.272	133.094	-247.178	(-252.53)	296.094	170.926	-125.168	(-112.362)
MDC Girls Share	.531	.494	-.037	(-71.901)	.482	.486	.005	(22.21)
MDC Fem.Teachers Share	.477	.405	-.072	(-73.767)	.531	.514	-.016	(-11.977)
Commuter to VS	.045	.173	.128	(55.238)	.03	.339	.308	(96.759)
Commuting to VS (km)	.132	.521	.389	(48.743)	.054	.996	.942	(55.032)
Commuting to RS (km)	3.045	6.412	3.368	(82.812)	1.909	3.915	2.006	(53.953)
Commuting to Gym (km)	2.604	6.383	3.779	(95.454)	2.672	5.12	2.448	(50.949)
Commuter	.566	.664	.098	(21.521)	.649	.71	.062	(11.168)
Observations	54465	15694			30245	10456		

Notes: In this table, we differentiate between control and treated students (between 15 and 20 years old) pre and post to the reform in 1969. A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. MDC = municipality-denomination-cohort, MG = multigrade, VS = Volksschule, RS = Realschule, Gym = Gymnasium, LM = labor market.

Source: Integrated dataset of Census 1970 and 1987 and Schools' Index 1964-1986. Own calculations.

Treatment Indicator

We determine each individual's likelihood for having been a multigrade student – considering each individual's municipality and denomination – computing two alternative continuous treatment indicators as explained in Section 5.4. The school-based indicator corresponds to the share of multigrade schools, the student-based indicator to the share of multigrade students of the respective municipality-denomination-cohort cell (MDC).[23] Table 5.4 shows that on average 26% of those students defined as control by the *binary* indicator are assigned a positive treatment probability by the *school-based* indicator. In contrast, 8.8% of those students defined as control by the *binary* indicator are assigned a positive treatment probability by the *student-based* indicator.

Controls

Using data from Saarland's Statistical Office, we obtain records on all primary and lower secondary schools from 1964 to 1986. Key figures like the numbers of male and female students and teachers, the number of classes, school's type, denomination and address are given for each school on an annual basis yielding more than 7500 school-year observations.[24] The school's address enables us to average over schooling conditions of schools of a given denomination in a given municipality in a given year. We then group the years into pre and post reform and match them to individuals in the baseline and follow up cohorts respectively via the municipality code while also considering an individual's denomination.[25] Importantly, for 80% of all schools (attended by roughly 50% of all students) a unique mapping between a student of a given denomination and the school of her denomination is possible (i.e. there is no need to match the student to an average of school characteristics of two or more schools of her denomination).

[23] See Table 5.3 for gaining an intuition of the different behavior of both indicators.

[24] We exclude special schools. Records for the years 1971/72 are missing completely. For 1966 one fifth of the data is missing but without region-specific missing patterns.

[25] In order to calculate average post-reform schooling conditions, we take schools' records from 1973-1986 into account. The cohorts of interest analyzed out of the 1987 Census are at most 20 years old in 1987 implying they entered primary school earliest in 1973.

By help of the schools' records we compute pre- and post-reform municipality - denomination - cohort (MDC) averages of class size, student-teacher ratio, school size (in terms of number of students), girls' share and female teachers' share. Table 5.4 compares the main schooling characteristics between schools in treated and control municipalities. Importantly, class size, the principal rivaling input when estimating the effect of multigrade schools, is a bit lower in treated regions (on average, there were 2.6 students less per class). Since a smaller class size has presumably beneficial effects on students' achievement, this fact will rather lead to underestimating the effects of the abolition of multigrade classes when not controlling for class size.

The census data provide us with a set of individual-level controls all displayed in Table 5.4, most of which are commonly used and self-explanatory. The differences between treated and control individuals are in line with expectations: Treated individuals are more likely to live in municipalities with less than 5,000 inhabitants (88% vs. 23%), and are more likely to have only one Volksschule (VS) in their municipality-denomination-cohort cell (MDC), namely by 90% vs. 30%. Moreover, treated individuals are less likely to be Catholic (70% vs. 80%).

Here we briefly discuss those controls with non-standard implications. In our setting, some standard controls like household size and marital status are potentially bad control because the reform likely affects marriage and/or fertility behavior (Lundborg *et al.*, 2012). The bad control case is even more pronounced for potential commuting costs. Students forced to commute are facing different effort costs than those attending school in direct vicinity. So continuing school at all is decided on altered premises. Simultaneously the implicit 'vicinity bonus' of lower secondary schools over higher education schools disappears in rural regions. Commuting anyway, ability-based school choice seems more natural than it has been with a *Volksschule* at walking distance and higher education schools at multiple kilometers' distance. Therefore we control for the distance to the nearest *Realschule* and/or *Gymnasium*. Importantly, however, we only include household size, marital status and commuting costs in an extended version of our regressions because we cannot rule out they are bad controls.

Sample Restrictions

Census data virtually cover all Saarlanders in each of the two survey years providing us with an unrestricted sample exceeding two million observations. We drop individuals younger than fifteen years because that is the minimum age for the outcomes we observe. Furthermore it is crucial to drop individuals between 21 and 32 years for two reasons.

First, before turning 21, people are still underage[26] such that their mobility is low. This matters because census data provide the municipality code of current residence and of school attendance. Fortunately, the residence-of-household definition ties children to their parents' address until they begin their own household.

Nevertheless, concerned with individuals moving reform-induced away from more rural regions (characterized by a higher likelihood of offering multigrade teaching) to urban regions we impose that underage restriction. It leaves us with a sample of main interest consisting of five consecutive birth cohorts with individuals who are between fifteen and twenty years old in either census. All of them attend primary and lower secondary school either strictly before or strictly after the reform takes place.

Second, although there is no panel structure at the individual level, observations of the 1970 Census reappear in the survey of 1987. Individuals between 32-37 years olds in 1987 have been past schooling age already in 1970 and are therefore untreated in either census. By construction their mobility cannot change reform-induced, so it is safe to include them as a control group. However the case is much more complicated for individuals between 21 and 32 years old in 1987. They have been partially treated because they are still in lower secondary school when the reform is rolled out in 1969. With respect to multigrade exposure they fall into a transition period with exceptional schooling conditions due to fundamental restructuring. Therefore, we exclude them from our sample. Note that the seventeen-year elapse between both censuses is just short enough to preclude that parents of the post-cohorts have already been treated. Otherwise multi-generational class composition effects could accumulate, a channel

[26] Legal definition as of 1970. For a subset of outcomes we run robustness checks restricting the sample to below 18 years, the legal threshold valid in 1987. This imitates what Lundborg *et al.* (2012) do facing the same problem.

established in Lundborg *et al.* (2012). Admittedly, the framework cannot rule out general equilibrium effects, a caveat that needs further investigation.

We furthermore restrict the sample to individuals for whom we have information on the outcomes of interest. In the end, our final dataset consists of 287,153 individuals when combining both age groups. When taking only into account the younger individuals of both censuses (aged between 15-20 years) the sample consists of 111,081 individuals.

5.6 Results

This section presents estimates of the impact of the abolition of multigrade schools on schooling and labor market outcomes. Our findings are in line with the literature suggesting a negative net effect from multigrade classes whenever other education inputs are not adapted accordingly. We show that results are robust to the inclusion of a wide range of individual characteristics and schooling covariates. Moreover we stratify the sample to investigate heterogeneity of the multigrade effect across subgroups. Throughout, we show (1) estimates of the DID estimation (i.e. not including the 32-37-year-olds as control group) using the school-based multigrade indicator, (2) estimates of the DDD estimation using the school-based multigrade indicator, (3) estimates of the DDD estimation using the student-based multigrade indicator. The multigrade indicators are calculated from the share of multigrade schools and multigrade students respectively. The latter respects the number of students (school size) upon averaging. Both indicators are continuously defined over 0 and 1 and measure each individual's multigrade exposure/treatment probability precisely.

Overall Results

Schooling Outcomes

Table 5.5 presents the main results based on the whole sample. We show estimates of the two different continuous DDD specifications (using the student-based multigrade indicator and the school-based multigrade indicator respectively) as well as estimates of a DID specification (using the school-based multigrade indicator). For each specification, we show estimates of the baseline approach (not including any controls), the core controls approach (not including potentially bad controls, see Section 5.5) and the extended controls approach (including all controls).[27] DID as well as DDD regression results displayed in Table 5.5 suggest that the abolition of denominational schools favorably influenced degree attainment. This finding is remarkably robust across our different specifications. According to the estimated coefficients the change from a multigrade school system to a single-grade school system significantly raised the average probability of attaining an intermediate secondary degree (*Mittlere Reife* or *Abitur*) by 7-11 percentage points (ppt), depending on the specification. The effect on having attained a high-school degree (*Abitur*) is also positive and indicates that the switch to a single-grade school system led to an increase of students holding a *Abitur* of around 5 ppt. A natural explanation of this finding would be that individuals spend more time on schooling because single-grade classes improve basic training. This in turn makes superior educational attainment accessible.

Professional Outcomes

The estimates in Table 5.5 show that the reform did not change the overall probability of being employed. Yet, we observe a reform-induced increase of the likelihood of holding a white-collar job and a reform-induced reduction of the likelihood of becoming a non-participant in the labor market (in other words, in the case of women, becoming a housewife). Interestingly, the labor market estimates get more precise

[27] We present the overall results for all three approaches. In the cases of the sensitivity analysis and the subgroup analysis, however, we only display the results of regressions including the core controls. The results of the other specifications are available upon request.

Table 5.5: Overall Effects on Schooling and Labour Market Outcomes of 15-20-Year-Olds

	Schooling		Labor Market			
	M.Reife/Abitur	Abitur	Employed	Blue-Collar	White-Collar	Non-Participant
DDD (student-based)						
Baseline	0.112***	0.0375	0.00277	-0.0165	0.0473	-0.0254
	[0.0367]	[0.0242]	[0.0395]	[0.0284]	[0.0311]	[0.0194]
Core Controls	0.112***	0.0407*	0.0187	0.00427	0.0463	-0.0457**
	[0.0363]	[0.0239]	[0.0446]	[0.0253]	[0.0304]	[0.0191]
Extended Controls	0.111***	0.0409*	0.0153	-0.00264	0.0463	-0.0391**
	[0.0367]	[0.0235]	[0.0423]	[0.0250]	[0.0287]	[0.0160]
DDD (school-based)						
Baseline	0.0903***	0.0520**	0.0296	-0.0285	0.0529**	-0.0194
	[0.0290]	[0.0229]	[0.0330]	[0.0241]	[0.0267]	[0.0171]
Core Controls	0.0898***	0.0534**	0.0371	-0.0162	0.0528**	-0.0320**
	[0.0286]	[0.0225]	[0.0348]	[0.0233]	[0.0262]	[0.0162]
Extended Controls	0.0912***	0.0529**	0.0396	-0.0193	0.0514**	-0.0280**
	[0.0287]	[0.0223]	[0.0333]	[0.0227]	[0.0242]	[0.0140]
DID (school-based)						
Baseline	0.0868***	0.0117	0.0515	-0.0242	0.0543*	-0.0164
	[0.0281]	[0.0122]	[0.0384]	[0.0271]	[0.0296]	[0.0130]
Core Controls	0.0823***	0.0115	0.0443	-0.00480	0.0358	-0.0187
	[0.0286]	[0.0124]	[0.0359]	[0.0264]	[0.0303]	[0.0116]
Extended Controls	0.0817***	0.0113	0.0324	-0.00750	0.0354	-0.0178**
	[0.0281]	[0.0121]	[0.0365]	[0.0261]	[0.0278]	[0.00794]
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
N (DDDstudent)	287153	287153	287153	287153	287153	287151
N (DDDschool)	287153	287153	287153	287153	287153	287151
N (DID)	111081	111081	111081	111081	111081	111079
Cluster (DDDstudent)	337	337	337	337	337	337
Cluster (DDDschool)	337	337	337	337	337	337
Cluster (DID)	333	333	333	333	333	333
Adj.R2 (DDDstudent)	0.129	0.0797	0.276	0.289	0.0931	0.510
Adj.R2 (DDDschool)	0.129	0.0797	0.276	0.289	0.0931	0.510
Adj.R2 (DID)	0.181	0.0660	0.172	0.234	0.189	0.550

Notes: This table shows in the upper part estimates of the DDD estimation using the student-based multigrade indicator, then it shows the estimates of the DDD estimation using the school-based multigrade indicator and in the bottom part it shows the estimates of the DID estimation (i.e. not including the 32-37-year-olds as control group) using the school-based multigrade indicator. The multigrade indicators are calculated from the share of multigrade schools and multigrade students respectively. For each specification, the table shows estimates of the baseline approach (not including any controls), estimates of the core controls approach (not including potentially bad controls) and estimates of the extended controls approach (including all controls).
Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.
Source: Integrated dataset of Census 1970 & 1987 and Schools' Index 1964-1986. Own calculations.

and larger when adding the control group of elder people, i.e. turning from the DID-estimation to the DDD-estimation. This indicates that the increased take-up of white-collar jobs is not due to a region-specific labor market trend. The global gain in white-collar employment seems to be partly driven by female labor market participation which is reflected in the housewife/non-labor-market-participation status declining by 3 ppt. Below, we discuss channels of gender-specific responsiveness to

the treatment in more detail. In sum, results suggest that reform-induced higher educational attainment led to an increase of better qualified employment.[28]

Sensitivity Analysis

Table 5.10 in the appendix shows the results of the main regressions – using the core controls approach – when restricting the sample in two different ways. For the sample used for regressions in the upper part of Table 5.10, we only take into account individuals for whom the municipality where they went to school is definitely known, i.e. we can exclude migration in order to take up employment elsewhere. This implies that this group of individuals represents a negative selection – they might be more afraid to move away from home or do not have sufficiently good skills to get employed elsewhere. The results in Table 5.10 are in line with this negative selection argument. While we observe a similar reaction to the switch from multigrade to single-grade teaching in terms of the attainment of a higher secondary degree, the labor market response is much smaller than for the whole sample. For results in the bottom part of Table 5.10, we restrict the sample to those individuals who live in those municipalities in which a unique mapping between individual and school is possible (since there is at maximum one school per denomination prior to the reform). This restriction makes a clean attribution of school controls possible. The disadvantage of this restriction is that we are left with the very small municipalities, and face, again, the problem of negative selection: those students staying in small villages are probably less ambitious. The results in Table 5.10 are very similar to the overall findings in Table 5.5. In contrast to the upper part of Table 5.10 we also find a significant negative effect on the likelihood to become a housewife/non-participant in the labor market.

[28] The importance to assess general equilibrium effects for policy recommendations is detailed in Heckman *et al.* (2014). As mentioned before the sizable period elapsing between pre- and post cohorts' outcomes heightens the probability that general equilibrium effects understate or overstate positive effects from improved education. Disentangling the partial effect we are interested in and the general effect offsetting it requires a joint estimation of skill supply and demand elasticity. The latter lies - for now - beyond the scope of our study.

Subgroup Analysis

Related studies motivate robustness checks by gender and denomination which we present in the following.

Boys & Girls

While the reasons for gender-specific reactions to education policies are still debated their existence has been shown repeatedly. Along these lines Angrist and Lavy (1999) find incentives pushing college certification rates only for Israeli girls. Deming *et al.* (2014) document gender-dependent attainment gains in US post-secondary education where only girls respond to higher school quality. These findings are complemented by relatively higher female responsiveness to tracking (Duflo *et al.*, 2011). However Whitmore (2005) draws on the STAR experiment to single out gender-neutral gains by class size reduction. As shown in Table 5.11 in the appendix, Saarland's data confirm girls' final grade attainment to improve more strongly than that of boys in the case of secondary education. While the switch from a multigrade system to a single-grade system led to a 11-16 ppt increase in a girl's likelihood to attain at least a secondary degree, it increased a boy's likelihood to attain such a degree by only 5-8 ppt which is already strong. Regarding the probability of attaining at least a high-school degree (*Abitur*), however, girls fare somewhat worse. Interestingly, as regards labor market outcomes, we do not observe large differences across gender and, moreover, the coefficients are not significant when splitting the sample. Yet, results in Table 5.11 show that the switch from a multigrade school system to a single-grade school system decreased the likelihood of becoming a housewife/non-participant in the labor market significantly for girls, but not for boys. What are potential explanations for girls benefiting more than boys from the disappearance of multigrade teaching? One possibility refers to girls being on average higher achieving than boys. Analogously it could be that their trajectories of improved education inputs are steeper. The literature also suggests girls to be less competitive than boys (Leuven and Rønning, 2014). Thus learning in highly heterogeneous multigrade groups might

be more demanding for them. Consequently, they profit more from the switch to single-grade classes.

Catholics & Protestants

Table 5.12 in the appendix shows the estimated coefficients for the sample stratified by denomination. Overall, it indicates that both groups of individuals benefited from the reform in terms of their educational outcomes. Surprisingly, Protestants seem to have gained by much more than Catholics did. Moreover, Table 5.12 shows insignificant and close-to-zero labor market effects for Catholics, while it indicates large and significant reform-induced gains for Protestants. What are potential explanations of this finding? Again, as in the case of explaining larger benefits of the reform accruing to girls than to boys, it could be that Protestant students are on average higher achieving than Catholic students and are therefore responding more to an increase of inputs into their education production function. This touches upon the challenging of the Weber Hypothesis (of Protestants' inherently superior *work ethics*) by Becker and Woessmann (2009) who connect wide-spread *literacy* to Protestants' prosperity. In a follow-up version of this paper, we will offer more evidence to gain a deeper understanding of the reasons for the heterogeneity of our findings with respect to denomination.

5.7 Conclusion and Outlook

This paper addresses the question how attending a multigrade school affects school attainment and labor market outcomes, and whether there are any differences by gender or denomination in this effect. To answer this question our analysis exploits the abolition of Saarland's denominational schools as a natural experiment that overcomes the main challenges of impact evaluations for policy design (McEwan, 2008).

The reform produces a sharp treatment effect, in terms of the variation of the reduced probability to attend a multigrade class caused by an exogenous event, namely the abolition of denominational schools. Based on a legal change that is rapidly and

comprehensively accomplished the setup provokes, if any, negligible anticipation or conditional-on-participation effects. Highly accurate school-level data allow us to control for rivaling changes in the educational infrastructure that are also implied by abandoning denominational tracking. The estimation approach based on triple differences plausibly identifies causal links between treatment and outcome candidates. Our results are remarkably robust across specifications and unambiguously suggest single-grade classes to be more beneficial for students' educational and labor market outcomes. Due to the reform treated students shift away from obtaining only a lower secondary degree (*Volksschulabschluss*) and a blue-collar job. Their probability to attain at least an intermediate secondary degree (*Realschulabschluss*) and to become a white-collar employee increases significantly when switching from a multigrade school system to a single-grade school system. Stratifying the main sample the emerging patterns line up with asymmetric treatment responses observed in related studies. Splitting the sample by denomination suggests that Protestant students profited more from the reform than Catholic students did. Moreover, we show that girls were more affected by the switch from a multigrade to a single-grade school system than boys. Our research approach provides external validity for the European context, which is particularly relevant in the light of the ongoing demographic change. To our knowledge this is the first study to exploit a large-scale experiment on multigrade classes in Germany. Policy interest in combination classes spans the globe but major empirical research is located in developing countries. Therefore, it suffers from limited external validity for the European context as third-world schooling bears many peculiarities. Saarland's data date back to the 1960s but the insights provided seem still more easily adaptable for use in Europe. The village schools we observe are much more likely to produce positive peer effects than schools in developing countries doomed by overage-for-grade students. Our findings nevertheless suggest that a beneficial multigrade system needs strategic adjustments. We conclude that peer effects based on student collaboration alone are no panacea which refutes the argument that reallocation is a *costless* way to improve education.

Still, there are some open questions that we want to address in a follow-up version of this paper: Why do we observe stronger effects of the reform for Protestants? So

far, we did not consider the pure effect of the abolition of *denominational* schools, but assume that the effects we find are the result of the disappearance of multigrade schools due to the abolition of denominational teaching. Yet, it might be that part of the *multigrade* effect is due to *denominational* teaching methods (that had a different impact in treated and control groups). Future research will thus try to disentangle the denominational effect from the multigrade effect. Furthermore, we will investigate in more depth why the shift from multigrade teaching to single-grade teaching has larger effects for girls. Using German data of the PIRLS study (Progress in International Reading Literacy Study) we will investigate whether gender-specific effects of multigrade teaching already arise at a young age. In particular, we will use the variation in the introduction of multigrade teaching in primary schools across German states between 2000 and 2010.

Appendix

Table 5.6: Mixed Grade Levels by Treated and Control Status of Catholic Students

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Mixed Levels/School	.986	5.571	4.585	(57.118)	.049	.226	.177	(10.436)
Not Mixing	.704	0	-.704	(-49.25)	.977	.834	-.143	(-16.189)
Mixing Two Levels	.1	.032	-.067	(-6.304)	.012	.107	.095	(13.586)
Mixing Three Levels	.048	.045	-.003	(-.296)	.006	.06	.053	(10.011)
Mixing Four Levels	.02	.072	.053	(6.119)	0	0	0	(-.572)
Mixing Five Levels	.027	.105	.078	(7.648)	.001	0	-.001	(-.991)
Mixing Six Levels	.03	.139	.109	(9.623)	.003	0	-.003	(-1.718)
Mixing Seven Levels	.038	.245	.207	(15.104)	0	0	0	(-.572)
Mixing Eight Levels	.025	.244	.218	(16.455)	0	0	0	(.)
Mixing All Levels	.008	.118	.109	(11.312)	0	0	0	(.)
Observations	1216	1021			2667	872		

Notes: A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. In this table, only Catholic students and the schools they attended are considered.
Source: Schools' Index 1964-1986. Own calculations.

Table 5.7: Mixed Grade Levels by Treated and Control Status of Protestant Students

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Mixed Levels/School	1.61	5.806	4.196	(29.635)	.087	.109	.023	(1.352)
Not Mixing	.58	0	-.58	(-24.854)	.94	.945	.005	(.571)
Mixing Two Levels	.078	.027	-.051	(-3.347)	.035	.034	-.001	(-.136)
Mixing Three Levels	.059	.038	-.021	(-1.402)	.024	.008	-.016	(-3.087)
Mixing Four Levels	.035	.056	.021	(1.431)	0	.001	.001	(1.673)
Mixing Five Levels	.08	.083	.002	(.124)	0	.003	.003	(2.901)
Mixing Six Levels	.067	.138	.072	(3.339)	.001	.008	.007	(3.513)
Mixing Seven Levels	.07	.252	.183	(7.165)	0	.001	.001	(1.673)
Mixing Eight Levels	.019	.25	.231	(9.919)	0	0	0	(.)
Mixing All Levels	.013	.156	.143	(7.302)	0	0	0	(.)
Observations	374	448			2607	932		

Notes: A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. In this table, only Protestant students and the schools they attended are considered.
Source: Schools' Index 1964-1986. Own calculations.

Table 5.8: Mixed Grade Levels by Treated and Control Status of Non-Denominational Students

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Mixed Levels/School	1.29	4.259	2.969	(8.423)	.071	.239	.168	(7.613)
Not Mixing	.623	0	-.623	(-9.717)	.95	.881	-.07	(-5.943)
Mixing Two Levels	.058	.121	.063	(1.25)	.031	.064	.033	(3.611)
Mixing Three Levels	.101	.121	.019	(.342)	.018	.031	.013	(1.89)
Mixing Four Levels	.043	.103	.06	(1.31)	0	.002	.002	(2.616)
Mixing Five Levels	.029	.155	.126	(2.563)	0	.007	.007	(4.54)
Mixing Six Levels	.101	.155	.054	(.905)	.001	.013	.012	(4.865)
Mixing Seven Levels	.029	.224	.195	(3.532)	0	.002	.002	(2.616)
Mixing Eight Levels	.014	.121	.106	(2.494)	0	0	0	(.)
Mixing All Levels	0	0	0	(.)	0	0	0	(.)
Observations	69	58			3087	452		

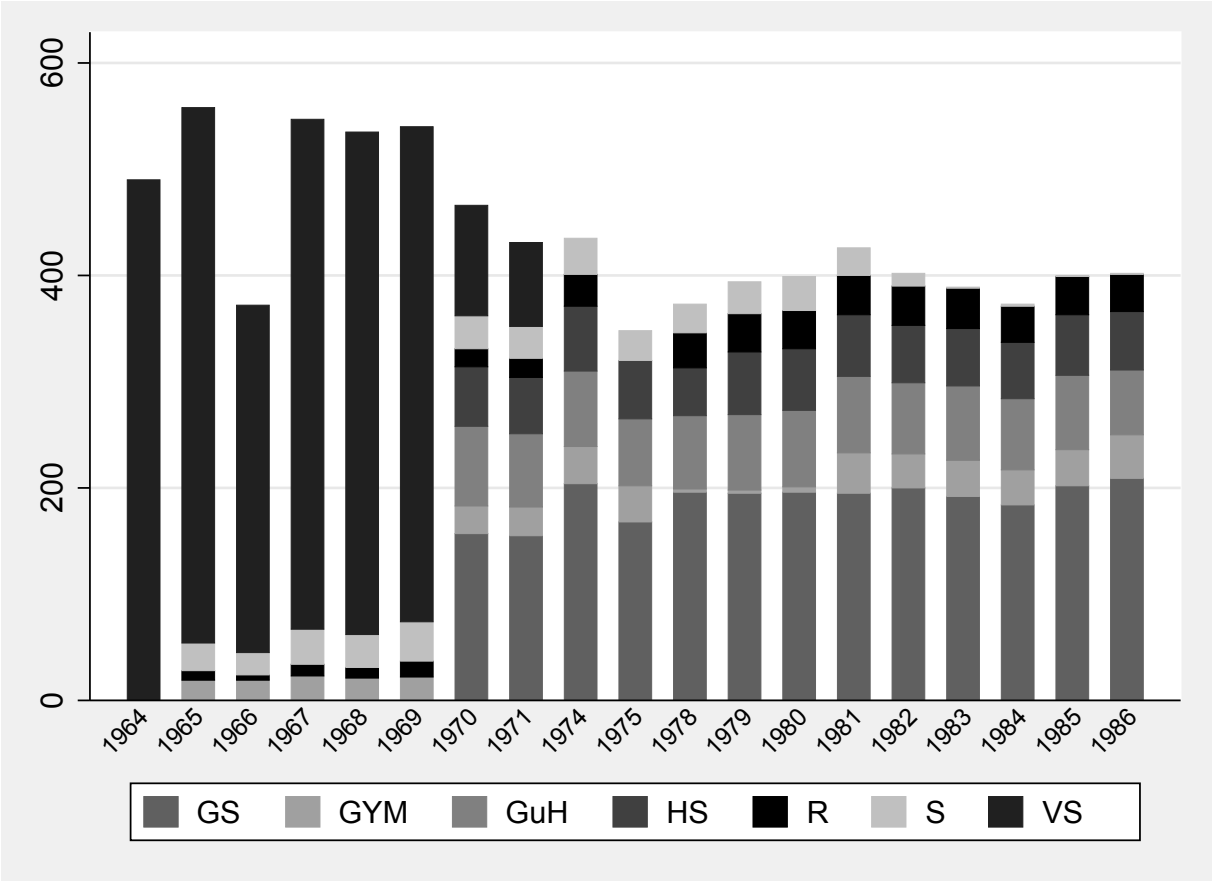
Notes: A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. In this table, only non-denominational students and the schools they attended are considered.
Source: Schools' Index 1964-1986. Own calculations.

Table 5.9: School Characteristics by Treated and Control Status of Non-Denominational Students

	PRE REFORM				POST REFORM			
	Control	Treated	Diff.	t-stat	Control	Treated	Diff.	t-stat
Class Size	31.087	31.397	.31	(.478)	23.017	21.788	-1.23	(-5.216)
Students/Teacher	31.043	32	.957	(1.377)	20.312	20.252	-.06	(-.223)
Students/School	321.826	149.81	-172.016	(-7.972)	251.168	210.701	-40.467	(-5.127)
Girls' Share	.493	.473	-.02	(-2.031)	.483	.483	0	(.052)
Female Teachers' Share	.574	.437	-.137	(-5.42)	.522	.559	.037	(3.807)
Teachers/School	10.188	4.655	-5.533	(-8.269)	12.539	10.564	-1.975	(-5.042)
Observations	69	58			3087	452		

Notes: A student is defined as *treated* if she is living in a municipality where all schools of her denomination were multigrade schools throughout all years prior to the reform in 1969. In this table, only non-denominational students and the schools they attended are considered.
Source: Schools' Index 1964-1986. Own calculations.

Figure 5.4: Main School Types' Distribution Over Time



Note: Schools' Index 1964-1986 (Own calculations). Records on 1972/73 and 1976/77 are missing completely. In 1964 only the type *Volksschule* (VS) is reported. 1966 about 20% of all types are missing. 1975 there are no records for Realschule (R) and 1978-80 for Gymnasiun (GYM). GS=Grundschule, GuH=Grund- und Hauptschule, HS=Hauptschule, S=Sonderschule.

Table 5.10: Effects on Schooling and Labour Market Outcomes – Alternative Sample Restrictions

	Schooling		**Labor Market**			
	M.Reife/Abitur	Abitur	Employed	Blue-Collar	White-Collar	Non-Participant
CERTAIN RESIDENCE						
DDD (student-based)						
Core Controls	0.106***	0.0284	-0.00849	-0.00316	0.0417	-0.0308
	[0.0342]	[0.0199]	[0.0535]	[0.0337]	[0.0482]	[0.0390]
DDD (school-based)						
Core Controls	0.0660**	0.0284	0.00131	0.0169	0.00293	-0.0153
	[0.0285]	[0.0183]	[0.0422]	[0.0235]	[0.0356]	[0.0293]
DID (school-based)						
Core Controls	0.0940***	0.00913	0.0589**	0.00972	-0.00211	-0.00702
	[0.0289]	[0.0116]	[0.0258]	[0.0306]	[0.0359]	[0.0190]
UNIQUE MAPPING						
DDD (student-based)						
Core Controls	0.0840**	0.0251	-0.000134	-0.00807	0.0548*	-0.0439**
	[0.0352]	[0.0221]	[0.0457]	[0.0263]	[0.0314]	[0.0201]
DDD (school-based)						
Core Controls	0.0529*	0.0210	0.0114	-0.00196	0.0413	-0.0355**
	[0.0276]	[0.0184]	[0.0371]	[0.0220]	[0.0258]	[0.0162]
DID (school-based)						
Core Controls	0.0652**	0.0136	0.0155	-0.00340	0.0272	-0.0161
	[0.0259]	[0.0118]	[0.0355]	[0.0280]	[0.0311]	[0.0109]
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
CERTAIN RESIDENCE						
N (DDDstudent)	132717	132717	132717	132717	132717	132716
N (DDDschool)	132717	132717	132717	132717	132717	132716
N (DID)	62445	62445	62445	62445	62445	62444
UNIQUE MAPPING						
N (DDDstudent)	125976	125976	125976	125976	125976	125975
N (DDDschool)	125976	125976	125976	125976	125976	125975
N (DID)	48836	48836	48836	48836	48836	48835

Notes: This table shows the results when restricting the sample in two different ways. In the upper part, only those individuals are taken into account for whom the municipality where they went to school is definitely known, i.e. we can exclude migration in order to take up employment elsewhere. In the bottom part, the sample is restricted to those individuals who live in those municipalities in which a unique mapping between individual and school is possible (since there is at maximum one school per denomination prior to the reform). In each part, first estimates of the DDD estimation using the student-based multigrade indicator are shown, then the estimates of the DDD estimation using the school-based multigrade indicator and then the estimates of the DID estimation (i.e. not including the 32-37-year-olds as control group) using the school-based multigrade indicator. The multigrade indicators are calculated from the share of multigrade schools and multigrade students respectively. For each specification, the estimates of the core controls approach (not including potentially bad controls) are shown.
Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.
Source: Integrated dataset of Census 1970 & 1987 and Schools' Index 1964-1986. Own calculations.

Table 5.11: Effects on Schooling and Labour Market Outcomes – Stratified by Gender

	Schooling		Labor Market			
	M.Reife/Abitur	Abitur	Employed	Blue-Collar	White-Collar	Non-Participant
BOYS						
DDD (student-based)						
Core Controls	0.0712**	0.0467	0.0236	-0.0286	0.0248	0.00828*
	[0.0355]	[0.0285]	[0.0465]	[0.0346]	[0.0328]	[0.00442]
DDD (school-based)						
Core Controls	0.0693**	0.0681**	0.0404	-0.0494	0.0486	0.00624
	[0.0318]	[0.0277]	[0.0361]	[0.0371]	[0.0358]	[0.00433]
DID (school-based)						
Core Controls	0.0554*	0.0135	0.0671*	0.00760	0.00457	-0.00439**
	[0.0326]	[0.0160]	[0.0407]	[0.0346]	[0.0353]	[0.00191]
GIRLS						
DDD (student-based)						
Core Controls	0.159***	0.0384	0.00721	0.0365	0.0655	-0.0953***
	[0.0469]	[0.0243]	[0.0689]	[0.0390]	[0.0534]	[0.0343]
DDD (school-based)						
Core Controls	0.116***	0.0411*	0.0304	0.0172	0.0542	-0.0666**
	[0.0390]	[0.0241]	[0.0500]	[0.0269]	[0.0422]	[0.0325]
DID (school-based)						
Core Controls	0.116***	0.00641	0.0179	-0.0202	0.0762*	-0.0378
	[0.0350]	[0.0144]	[0.0511]	[0.0369]	[0.0452]	[0.0232]
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
BOYS						
N (DDDstudent)	146633	146633	146633	146633	146633	146631
N (DDDschool)	146633	146633	146633	146633	146633	146631
N (DID)	58042	58042	58042	58042	58042	58040
GIRLS						
N (DDDstudent)	140520	140520	140520	140520	140520	140520
N (DDDschool)	140520	140520	140520	140520	140520	140520
N (DID)	53039	53039	53039	53039	53039	53039

Notes: This table shows the results when stratifying the sample by gender. For each subgroup, first estimates of the DDD estimation using the student-based multigrade indicator are shown, then the estimates of the DDD estimation using the school-based multigrade indicator and then the estimates of the DID estimation (i.e. not including the 32-37-year-olds as control group) using the school-based multigrade indicator. The multigrade indicators are calculated from the share of multigrade schools and multigrade students respectively. For each specification, the estimates of the core controls approach (not including potentially bad controls) are shown.
Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.
Source: Integrated dataset of Census 1970 & 1987 and Schools' Index 1964-1986. Own calculations.

Table 5.12: Effects on Schooling and Labour Market Outcomes – Stratified by Denomination

	Schooling		**Labor Market**			
	M.Reife/Abitur	Abitur	Employed	Blue-Collar	White-Collar	Non-Participant
CATHOLICS						
DDD (student-based)						
Core Controls	0.0726**	0.0239	0.0132	0.0158	0.0120	-0.0292
	[0.0360]	[0.0244]	[0.0424]	[0.0277]	[0.0335]	[0.0199]
DDD (school-based)						
Core Controls	0.0658**	0.0354*	0.0328	-0.0000945	0.0192	-0.0173
	[0.0274]	[0.0202]	[0.0330]	[0.0238]	[0.0273]	[0.0164]
DID (school-based)						
Core Controls	0.0800***	0.00828	0.0430	-0.00836	0.0352	-0.0142
	[0.0282]	[0.0126]	[0.0358]	[0.0271]	[0.0304]	[0.0104]
PROTESTANTS						
DDD (student-based)						
Core Controls	0.299***	0.0335	-0.216	-0.0993	0.150	0.0131
	[0.0888]	[0.0543]	[0.174]	[0.0818]	[0.103]	[0.0664]
DDD (school-based)						
Core Controls	0.110*	0.0735	-0.0744	-0.108*	0.138**	0.00421
	[0.0588]	[0.0553]	[0.0985]	[0.0617]	[0.0609]	[0.0371]
DID (school-based)						
Core Controls	0.0332	0.0328	0.138	0.0658	-0.0303	-0.0366
	[0.0836]	[0.0386]	[0.114]	[0.0684]	[0.0796]	[0.0390]
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
CATHOLICS						
N (DDDstudent)	217373	217373	217373	217373	217373	217371
N (DDDschool)	217373	217373	217373	217373	217373	217371
N (DID)	86288	86288	86288	86288	86288	86286
PROTESTANTS						
N (DDDstudent)	61070	61070	61070	61070	61070	61070
N (DDDschool)	61070	61070	61070	61070	61070	61070
N (DID)	22837	22837	22837	22837	22837	22837

Notes: This table shows the results when stratifying the sample by denomination. For each subgroup, first estimates of the DDD estimation using the student-based multigrade indicator are shown, then the estimates of the DDD estimation using the school-based multigrade indicator and then the estimates of the DID estimation (i.e. not including the 32-37-year-olds as control group) using the school-based multigrade indicator. The multigrade indicators are calculated from the share of multigrade schools and multigrade students respectively. For each specification, the estimates of the core controls approach (not including potentially bad controls) are shown.
Standard errors are clustered at the municipality level and are shown in parenthesis: *p < 0.10, ** p<0.05, ***p<0.010.
Source: Integrated dataset of Census 1970 & 1987 and Schools' Index 1964-1986. Own calculations.

Bibliography

Abramitzky, R. and Lavy, V. (2011). How Responsive is Investment in Schooling to Changes in Redistribution Policies and in Returns. *Econometrica*, **82**(4), 1241–1272.

Aizer, A. (2004). Home Alone: Supervision after School and Child Behavior. *Journal of Public Economics*, **88**(9-10), 1835–1848.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, **84**(3), 488–500.

Almond, D. and Currie, J. (2011). *Human Capital Development before Age Five*, volume 4 of *Handbook of Labor Economics*, chapter 15, pages 1315–1486. Elsevier.

Angrist, J. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, J. D. (2004). American Education Research Changes Track. *Oxford Review of Economic Policy*, **20**(2), 198–212.

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, **114**(2), 533–575.

Aschenbach, T. M. and Edelbrock, C. (1981). Behavior Problems and Competencies. Reported by Parents of Normal and Disturbed Children aged 4 through 16. *Monographs of the Society for Research in Child Development*, **46**(188).

Baker, M., Gruber, J., and Milligan, K. (2008). Universal Child Care, Maternal Labor Supply, and Family Well-Being. *Journal of Political Economy*, **116**(4), 709–745.

Bastos, P. and Cristia, J. (2012). Supply and Quality Choices in Private Child Care Markets: Evidence from Sao Paulo. *Journal of Development Economics*, **98**(2), 242 – 255.

Bauernschuster, S. and Schlotter, M. (2015). Public Child Care and Mothers' Labor Supply. Evidence from two Quasi-Experiments. *Journal of Public Economics*, **123**, 1 – 16.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., and Weiss, M. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske u. Budrich.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition*. Chicago: University of Chicago Press.

Becker, S. O. and Woessmann, L. (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *The Quarterly Journal of Economics*, **124**(2), 531–596.

Berkemeyer, N., Bos, W., Manitius, V., Hermstein, B., Bonitz, M., and Semper, I. (2014). *Chancenspiegel 2014: Regionale Disparitäten in der Chancengerechtigkeit und Leistungsfähigkeit der deutschen Schulsysteme*. Verlag Bertelsmann Stiftung.

Berlinski, S., Galiani, S., and Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, **93**(1-2), 219–234.

Bernal, R. and Keane, M. P. (2011). Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers. *Journal of Labor Economics*, **29**(3), 459–512.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *American Economic Review*, **95**(1), 437–449.

Blau, D. (1999). The Effect Of Income On Child Development. *The Review of Economics and Statistics*, **81**(2), 261–276.

Blau, D. and Currie, J. (2006). *Pre-School, Day Care, and After-School Care: Who's Minding the Kids?*, volume 2 of *Handbook of the Economics of Education*, chapter 20, pages 1163–1278. Elsevier.

Blau, D. M. (1997). The Production of Quality in Child Care Centers. *The Journal of Human Resources*, **32**(2), 354–387.

Blau, D. M. (2007). Unintended Consequences of Child Care Regulations. *Labour Economics*, **14**(3), 513–538.

Blau, D. M. and Hagy, A. P. (1998). The Demand for Quality in Child Care. *Journal of Political Economy*, **106**(1), 104–146.

BMFSFJ (2014). *Vierter Zwischenbericht zur Evaluation des Kinderförderungsgesetzes*. Bundesministerium für Familie, Senioren, Frauen und Jugend.

Bowlby, J. (1969). *Attachment and Loss*. Hogarth, London.

Bowles, S., Gintis, H., and Osborne, M. (2001). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature*, **39**(4), 1137–1176.

Brooks, P., Mojica, C., and Land, R. (1995). Final Evaluation Report. Longitudinal Study of LA's Best After School Education and Enrichment Program. Report, Center for the Study of Evaluation. University of California, Los Angeles.

Brooks-Gunn, J., Han, W.-J., and Waldfogel, J. (2002). Maternal employment and child cognitive outcomes in the first three years of life: The nichd study of early child care. *Child Development*, **73**(4), 1052–1072.

Cannon, J., Jacknowitz, A., and Painter, G. (2006). Is Full Better than Half? Examining the Longitudinal Effects of Full-Day Kindergarten Attendance. *Journal of Policy Analysis and Management*, **25**(2), 299–321.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, **81**(3), 855–882.

Cascio, E. (2009). Do Investments in Universal Early Education Pay Off? Long-Term Effects of Introducing Kindergartens into Public Schools. Technical Report 14951, NBER.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review*, **92**(5), 1308–1334.

Chamorro-Premuzic, T. and Furnham, A. (2005). *Personality and Intellectual Competence*. CMahwah, NJ: Lawrence Erlbaum.

Chamorro-Premuzic, T. and Furnham, A. (2006). Intellectual Competence and the Intelligent Personality: A Third Way in Differential Psychology. *Review of General Psychology*, **10**(3), 251–267.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014a). Is the United States still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review*, **104**(5), 141–147.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, **104**(9), 2633–2679.

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

Coleman, J. S. (1968). Equality of Educational Opportunity Study (EEOS). *Equity & Excellence in Education*, **6**(5), 19–28.

Cornelissen, T., Dustmann, C., Raute, A., and Schoenberg, U. (2016). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. mimeo, University College London, Department of Economics.

Cortes, K. E. and Goodman, J. S. (2014). Ability-Tracking, Instructional Time, and Better Pedagogy: The Effect of Double-Dose Algebra on Student Achievement. *The American Economic Review*, **104**(5), 400–405.

Cryan, J., Sheehan, R., Wiechel, J., and I, B.-H. (1992). Success Outcomes of Full-Day Kindergarten: More Positive Behavior and Increased Achievement in the Years After. *Early Childhood Research Quarterly*, **7**, 187–203.

Cullen, J. B., Jacob, B. A., and Levitt, S. (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica*, **74**(5), 1191–1230.

Cunha, F., Heckman, J. J., and Lochner, L. (2006). *Interpreting the Evidence on Life Cycle Skill Formation*, volume 1 of *Handbook of the Economics of Education*, chapter 12, pages 697–812. Elsevier.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, **47**(1), 87–122.

Dahl, G. B. and Lochner, L. (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *The American Economic Review*, **102**(5), 1927–1956.

Datta-Gupta, N. and Simonsen, M. (2010). Non-Cognitive Child Outcomes and Universal High Quality Child Care. *Journal of Public Economics*, **94**, 30–43.

Datta-Gupta, N. and Simonsen, M. (2015). Academic Performance and Type of Early Childhood Care. IZA Discussion Papers 9045, Institute for the Study of Labor (IZA).

Davis-Kean, P. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, **19**(2), 294–304.

DeCicca, P. (2007). Does Full-Day Kindergarten matter? Evidence from the First Two Years of Schooling. *Economics of Education Review*, **26**, 67–82.

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School Choice, School Quality, and Postsecondary Attainment. *The American Economic Review*, **104**(3), 991–1013.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, **100**(3), 1238–60.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review*, **101**(5), 1739–1774.

Duncan, G., Dowsett, C., Claessens, A., Magnuson, K., Huston, A., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., and Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, **43**(6), 1428–46.

Duncan, G. J. and Gibson-Davis, C. M. (2006). Connecting Child Care Quality to Child Outcomes. *Evaluation Review*, **30**(5), 611–630.

Dustmann, C., Raute, A., and Schönberg, U. (2012). Does Universal Child Care Matter? Evidence from a Large Expansion in Pre-School Education. *mimeo*.

EENEE (2015). The Impact of School Size and Consolidations on Quality and Equity in Education. Technical Report 19, EENEE (European Expert Network on Economics of Education).

Felfe, C. and Hsin, A. (2012). Maternal Work Conditions and Child Development. *Economics of Education Review*, **31**(6), 1037–1057.

Felfe, C. and Lalive, R. (2014). Does Early Child Care Help or Hurt Child Development? Discussion paper 2014:22, IFAU.

Felfe, C. and Zierow, L. (2014). After-School Center-Based Care and Children's Development. *B.E. Journal of Economic Analysis and Policy (Contributions)*, **14**(4), 1299–1336.

Felfe, C., Nollenberger, N., and Rodríguez-Planas, N. (2015). Can't buy mommy's love? Universal Childcare and Children's Long-term Cognitive Development. *Journal of Population Economics*, **28**(2), 393–422.

Fitzpatrick, M. D. (2008). Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement. *The B.E. Journal of Economic Analysis & Policy*, **8**(1), 46.

Friesen, J., Krauth, B., and Sattari, R. (2013). Full-day Kindergarten and the Development of Non-Cognitive Skills. mimeo, Simon Fraser University.

Galor, O. (2005). Chapter 4: From Stagnation to Growth: Unified Growth Theory. In P. Aghion and S. N. Durlauf, editors, *Handbook of Economic Growth*, volume 1, Part A of *Handbook of Economic Growth*, pages 171 – 293. Elsevier.

Glewwe, P., Kremer, M., and Moulin, S. (2009). Many Children Left Behind? Textbooks and Test Scores in Kenya. *American Economic Journal: Applied Economics*, **1**(1), 112–135.

Gonzalez, L. (2013). The Effects of a Universal Child Benefit. *American Economic Journal: Economic Policy*, **5**(3).

Goodman, R. N. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, **5**(38), 581–586.

Gormley Jr., W. T., Phillips, D., and Gayer, T. (2008). THE EARLY YEARS: Preschool Programs Can Boost School Readiness. *Science*, **320**(5884), 1723–1724.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2016). Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment, and Married Female Labor-Force Participation. *American Economic Journal: Macroeconomics*, **8**(1), 1–41.

Gregg, P. and Machin, S. (1999). Childhood Disadvantage and Success or Failure in the Labour Market. In D. Blanchflower and R. Freeman, editors, *Youth Employment and Joblessness in Advanced Countries*, Cambridge, MA. National Bureau of Economic Research.

Gregg, P. and Machin, S. (2001). The Relationship Between Childhood Experiences, Subsequent Educational Attainment and Adult Labour Market Performance. In

K. Vleminckx and T. Smeeding, editors, *Child Well Being in Modern Nations: What do we Know?*

Grissmer, D., Grimm, K., Aiyer, S., Murrah, W., and Steele, J. (2010). Fine Motor Skills and Early Comprehension of the World: Two New School Readiness Indicators. *Developmental Psychology*, **46**(5), 1008–17.

Grossman, J., Price, M., Fellerath, V., Jucovy, L., Kotloff, L., Raley, R., and Walker, K. (2002). Multiple Choices After School: Findings from the Extended-Service Schools Initiative. Report, Public/Private Ventures, Philadelphia.

Gullo, D. (2000). The Long-Term Educational Effects of Half-Day vs. Full-Day Kindergarten. *Early Childhood Development and Care*, **160**, 17–24.

Han, W. (2005). Maternal Nonstandard Work Schedules and Child Cognitive Outcomes. *Child Development*, **76**, 137–154.

Hanushek, E., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, **18**(5), 527–544.

Hanushek, E. A. and Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, **46**(3), 607–68.

Hart, B. and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes, Baltimore, MD.

Hattie, J. A. (2002). Classroom Composition and Peer Effects. *International Journal of Educational Research*, **37**(5), 449–481.

Havnes, T. and Mogstad, M. (2011). No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes. *American Economic Journal: Economic Policy*, **3**(2), 97–129.

Heckman, J., Lochner, L., and Taber, C. (2014). General Equilibrium Effects of Schooling Policy. Technical report.

Heckman, J. J. (2011). The Economics of Inequality – The Value of Early Childhood Education. *American Educator*, **35**(1), 31–35.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006a). *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, volume 1 of *Handbook of the Economics of Education*, chapter 7, pages 307–458. Elsevier.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006b). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, **24**(3), 411–482.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, **94**(1-2), 114–128.

Heimer, A., Henkel, M., and Donges, D. (2011). Vereinbarkeit von Familie und Beruf mit Schulkindern. Report, Federal Ministry of Family, Senior Citizens, Women and Youth, Germany.

Hill, J., Waldfogel, J., Brooks-Gunn, J., and Han, W.-J. (2005). Maternal Employment and Child Development: A Fresh Look Using Newer Methods. *Development Psychology*, **41**(6), 833–850.

Hotz, V. J. and Xiao, M. (2011). The Impact of Regulations on the Supply and Quality of Care in Child Care Markets. *American Economic Review*, **101**(5), 1775–1805.

Hsin, A. and Felfe, C. (2014). When Does Time Matter? Maternal Employment, Children's Time With Parents, and Child Development. *Demography*, **51**(5), 1867–1894.

Huang, D., Gribbons, B., Kim, K., Lee, C., and Baker, E. (2000). A decade of results: The impact of LA's Best after school enrichment program on subsequent student achievement and performance. Report, UCLA Center for the Study of Evaluation, Los Angeles, CA.

Huang, D., Leon, S., La Torre, D., and Mostafavi, S. (2008). Examining the relationship between LA's Best program attendance and academic achievement of LA's Best students. Report, CRESST Report 749.

Huesken, K. (2010). Kita vor Ort: Betreuungsatlas auf Ebene der Jugendamtsbezirke 2010. Technical report, Deutsches Jugendinstitut e.V. und Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik.

Hüsken, K. and Riedel, B. (2012). Kommunen im Endspurt - Zum Stand des Ausbaus der Kindertagesbetreuung für unter Dreijährige. *DJI Impulse*, **02**, 4–7.

Iversen, J. M. and Bonesrønning, H. (2015). Conditional Gender Peer Effects? *Journal of Behavioral and Experimental Economics*, **55**, 19–28.

Jacob, B. A., Lefgren, L., and Sims, D. P. (2010). The Persistence of Teacher-Induced Learning. *Journal of Human Resources*, **45**(4), 915–943.

Janet Currie, R. H. (1999). Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight. *The American Economic Review*, **89**(2), 245–250.

Johansson, E.-A. and Lindahl, E. (2008). The Effects of Mixed-Age Classes in Sweden. *IFAU Working Paper Series*, **21**.

Johnson, C. J., Boller, K., Young, M., Thomas, J., and Gonzalez, D. (2015). A Closer Look: Informal Child Care Arrangements and Support in California (Issue Brief). Technical report, Mathematica Policy Research.

Johnson, R., Kalil, A., and Dunifon, R. (2005). Employment Patterns of Less-Skilled Workers: Links to Children's Behavior and Academic Progress. *Demography*, **49**, 747–772.

Jones, S. (2013). Class size versus class composition: What matters for learning in East Africa? *WIDER Working Paper Series*, **065**.

Kamette, F. (2011). Organisation of School Time in the European Union. Report, Fondation Robert Schuman, European issues no 212.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, **75**(1), 83–119.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, **114**(2), 497–532.

Lange, J., Riedel, B., Fuchs-Riechlin, K., Schilling, M., and Leu, H. R. (2007). Zahlenspiegel: Kindertagesbetreuung im Spiegel der Statistik. Technical report, Deutsches Jugendinstitut e.V. und Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik.

Lavy, V., Paserman, M. D., and Schlosser, A. (2012). Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *Economic Journal*, **122**(559), 208–237.

Leuven, E. and Rønning, M. (2014). Classroom Grade Composition and Pupil Achievement. *The Economic Journal*, pages 1–39.

Lichtblau, K. (2009). 50 Jahre Saarland: Wirtschaft Saarland 1959 bis 2009 Wie hat sich das Saarland in den letzten 50 Jahren entwickelt - ein Bundesländervergleich. Technical report, Institut der deutschen Wirtschaft Köln (IW Consult GmbH), Köln.

Linver, M., Brooks-Gunn, J., and Kohen, D. (2002). Family Processes as Pathways from Income to Young Children's Development. *Developmental Psychology*, **38**, 719–734.

Little, A. W. (2001). Multigrade Teaching: Towards an International Research and Policy Agenda. *International Journal of Educational Development*, **21**(6), 481–497.

Lundborg, P., Nilsson, A., and Rooth, D.-O. (2012). Parental Education and Offspring Outcomes: Evidence from the Swedish Compulsory Schooling Reform. *IZA Discussion Paper Series*, **6570**.

Magnuson, K. and Duncan, G. J. (2016). Can Early Childhood Interventions Decrease Inequality of Economic Opportunity? *RSF*, **2**(2), 123–141.

Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, **26**(1), 33 – 51. The Economics of Early Childhood Education.

Mason, D. A. and Burns, R. B. (1996). 'Simply No Worse and Simply No Better' May Simply Be Wrong: A Critique of Veenman's Conclusion About Multigrade Classes. *Review of Educational Research*, **66**(3), 307–322.

McEwan, P. J. (2008). Evaluating Multigrade School Reform in Latin America. *Comparative Education*, **44**(4), 465–483.

McLeod, J. D. and Kaiser, K. (2004). Childhood Emotional and Behavioral Problems and Educational Attainment. *American Sociological Review*, **69**(5), 636–658.

McLoyd, V. (1998). Socioeconomic Disadvantage and Child Development. *American Psychologist*, **53**, 185–204.

Mercer, J. (2006). *Understanding Attachment: Parenting, Child Care, and Emotional Development*. Praeger, Westport (Connecticut).

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., and Caspi, A. (2011). A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety. *Proceedings of the National Academy of Sciences*, **108**(7), 2693–2698.

Mulkeen, A. and Higgings, C. (2009). Multigrade Teaching in Sub-Saharan Africa. *World Bank Working Paper Series*, **173**.

Müller, K.-U., Tsiasioti, C., Spiess, C. K., Wrohlich, K., Bügelmayer, E., Haywood, L., Peter, F., Ringmann, M., and Witzke, S. (2013). Förderung und Wohlergehen von Kindern. Politikberatung kompakt Nr.73. Technical report, DIW Berlin.

Noboa Hidalgo, G. and Urzua, S. (2012). The Effect of Participation in Public Childcare Centers: Evidence from Chile. *Journal of Human Capital*, **6**(1).

OECD (2011). *Doing Better for Families*. Paris: OECD Publishing.

OECD (2012). *Education at a Glance 2012: OECD Indicators*. Paris: OECD Publishing.

OECD (2015). *Immigrant Students at School – Easing the Journey towards Integration*. Paris: OECD Publishing.

OECD (2016). *Education at a Glance*. Paris: OECD Publishing.

Patall, E., Cooper, H., and Allen, A. B. (2010). Extending the School Day or School Year: A Systematic Review of Research (1985 -2009). *Review of Educational Research*, **80**(3), 401–436.

Piketty, T. (2014). *Capital in the 21st century*. Harvard University Press.

Pischke, J. and Wachter, T. V. (2005). Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation. *The Review of Economics and Statistics*, **90(3)**, 592–598.

Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update. *World Development*, **22**(9), 1325–1343.

Rathburn, A. and West, J. (2004). From Kindergarten through Third Grade: Children's Beginning School Experiences. *Washington, D.C.: National Center for Education Statistics*.

Raver, C. (2003). Does work pay psychologically as well as economically? The role of employment in predicting depressive symptoms and parenting among low-income families. *Child Development*, **74**, 1720–36.

Riedel, B., Gadow, T., van Santen, E., Fuchs, K., Schilling, M., and Leu, H. R. (2005). Zahlenspiegel 2005: Kindertagesbetreuung im Spiegel der Statistik. Technical report, Deutsches Jugendinstitut e.V. und Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik.

Rothstein, J. M. (2006). Good principals or good peers? Parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *The American Economic Review*, **96**(4), 1333–1350.

Russell, J. V., Rowe, K. J., and Hill, P. W. (1998). Effects of Multigrade Classes on Student Progress in Literacy and Numeracy: Quantitative Evidence and Perceptions of Teachers and School Leaders. In *Annual Meeting of the Australian Association for Research in Education*, Adelaide.

Schilling, M. (2009). Zwischen konfessionellen Trägern und Wirtschaftsunternehmen - stabile Trägerlandschaften. *Komdat, Kommentierte Daten der Kinder- und Jugendhilfe*, **12**, 16–17.

Schnock, B. (2009). Eignung von Tagespflegepersonen in der Kindertagespflege. Technical report, Deutsches Jugendinstitut e.V.

Schober, P. S. and Spieß, C. K. (2014). Local Day-Care Quality and Maternal Employment: Evidence from East and West Germany. SOEPpapers on Multidisciplinary Panel Data Research, DIW. no. 649.

Schwandt, H. and Wuppermann, A. (2016). The youngest get the pill: ADHD misdiagnosis in germany, its regional correlates and international comparison. *Labour Economics*, **43**, 72 – 86.

Sell, S. and Kukula, N. (2011). Vergütung der Kindertagespflege: Bestandsaufnahme und Modelle einer leistungsorientierten Vergütungssystematik. Technical report, Institut für Bildungs- und Sozialpolitik der Hochschule Koblenz (ibus).

Simonsen, M. (2005). Availability and Price of High Quality Day Care and Female Employment. Economics working papers, School of Economics and Management, University of Aarhus.

Smith, A. (1904). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edwin Cannan, Library of Economics and Liberty. Retrieved November 19, 2016 from the World Wide Web: http://www.econlib.org/library/Smith/smWN.html.

Stephens, M. and Yang, D.-Y. (2014). Compulsory Education and the Benefits of Schooling. *American Economic Review*, **104**(6), 1777–1792.

Strunz, E. (2011). Kita vor Ort: Betreuungsatlas auf Ebene der Jugendamtsbezirke 2011. Technical report, Deutsches Jugendinstitut e.V. und Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik.

Sunde, U. and Vischer, T. (2015). Human Capital and Growth: Specification Matters. *Economica*, **82**(326), 368–390.

Thomas, J. (2012). Combination Classes and Educational Achievement. *Economics of Education Review*, **31**(6), 1058–1066.

Tietze, W., Becker-Stoll, F., Bensel, J., Eckhardt, A. G., Haug-Schnabel, G., Kalicki, B., Keller, H., and Leyendecker, B. (2012). Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit - Fragestellungen und Ergebnisse im Überblick. Technical report, Hrsg.

Todd, P. E. and Wolpin, K. I. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, **1**(1), 91–136.

Veenman, S. (1995). Cognitive and Noncognitive Effects of Multigrade and Multi-Age Classes: A Best-Evidence Synthesis. *Review of Educational Research*, **65**(4), 319–381.

Vivalt, E. (2015). Heterogeneous Treatment Effects in Impact Evaluation. *American Economic Review*.

Wahl, K. (2008). Lachen, weinen, ärgern: Die Gefühlswelt der Kinder - Emotionen, Kompetenzen, Risiken. Deskription der Daten der ersten Welle des DJI Kinderpanel. Technical report, Deutsches Jugendinstitut e.V.

Waldfogel, J., Han, W.-J., and Brooks-Gunn, J. (2002). The Effects of Early Maternal Employment on Child Cognitive Development. *Demography*, **39**(2), 369–392.

Walson, J. and West, J. (2004). Full-day and half-day Kindergarten in the United States: findings from the Early Childhood Longitudinal Study, Kindergarten class of 1998-99. *U.S.Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office*.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *The American Economic Review*, **95**(2), 199–203.

Windle, M. and Lerner, R. (1986). Reassessing the Dimensions of Temperamental Individuality across the Life span. The Revised Dimensions of Temperamant Survey (DOTS-R). *Journal of Adolescent Research*, **1**, 213–230.

Woessmann, L. (2016). The Economic Case for Education. *Education Economics*, **24**(1), 3–32.

Yeung, J., Linver, M., and Brooks-Gunn, J. (2002). How Money Matters for Young Children's Development: Parental Investment and Family Processes. *Child Development*, **73**, 1861–1879.

Curriculum Vitae

Larissa Zierow
born on September 2 1985 in Berlin, Germany

Since April 2017	Economist at the Ifo Institute Leibniz Institute for Economic Research at the University of Munich, Germany
April 2012 – March 2017	PhD Student, Munich Graduate School of Economics at the Ludwig-Maximilians University Munich, Germany
April 2012 – March 2017	Research Assistant Seminar for Population Economics at the Ludwig-Maximilians University Munich, Germany
Dec. 2011 – March 2012	Research Assistant Swiss Institute for Empirical Economic Research at the University of St. Gallen, Switzerland
June 2011 – Nov. 2011	Research Assistant Institute of Economics at the University of St. Gallen, Switzerland
Aug. 2010 – Dec. 2010	Exchange semester at the Claremont Graduate University, CA, USA
Sep. 2009 – Aug. 2011	Master of Arts in Economics at the University of St. Gallen, Switzerland
Feb. 2008 – July 2008	Exchange semester at the University of Bologna, Italy
Sep. 2006 – Aug. 2009	Bachelor of Arts in Business Economics at the University of Witten/Herdecke, Germany
June 2005	Abitur Goethe-Gymnasium Berlin-Wilmersdorf, Germany