CESifo WORKING PAPERS

8618
2020

Original Version: October 2020
This Version: May 2022

The Impact of All-Day Schools on Student Achievement – Evidence from Extending School Days in German Primary Schools

Arnim Seidlitz, Larissa Zierow

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8618*

The Impact of All-Day Schools on Student Achievement – Evidence from Extending School Days in German Primary Schools

Abstract

This paper studies the effect of longer school days − induced by voluntary all-day programs in German primary schools − on school performance. Facing the challenge of selection into all-day school programs, we instrument all-day school expansion with construction subsidies from a large federal investment project. We combine data from the representative National Educational Panel Study covering more than 5'000 primary school students with municipality-level information on federal subsidies. Results show that all-day programs lead to improvements in language and math grades and to a higher probability of attending the academic track after primary school. Heterogeneity analysis suggests that the programs do not reduce educational inequality.

JEL-Codes: J130, I280, I240.

Keywords: all-day school, skill development, educational inequality.

Arnim Seidlitz
Humboldt University Berlin &
Berlin School of Economics
Unter den Linden 6
Germany – 10099 Berlin
arnim.seidlitz@hu-berlin.de

Larissa Zierow
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5
Germany – 81679 Munich
zierow@ifo.de

April 2022
We thank Eric Hanushek, Ludger Woessmann, Bernd Fitzenberger and seminar participants at HU Berlin, ifo Institue, the 4th international NEPS Conference 2019, the UCL-IAB workshop 2020, the EALE-SOLE-AASLE World Conference 2020, the EEA Congress 2020, and the CESifo Economics of Education Conference 2020 for their valuable input. Financial support by NORFACE through the project "The impact of childhood circumstances on individual outcomes over the life-course – IMCHILD" as well as by the project "Efficiency and Equity in Education: Quasi-Experimental Evidence from School Reforms across German States (EffEE)" generously funded by the Leibniz Association under its competitive procedure, is gratefully acknowledged. This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 2, survey waves 3 to7. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. All remaining errors are our own.

1 Introduction

All-day schools and school-based extracurricular afternoon activities are often considered to be a vehicle for providing favorable learning conditions and for increasing equality of opportunity in the education system. Increased time spent at school can provide extended learning opportunities and improved individual support to foster children's academic and psychosocial development, especially those from more disadvantaged backgrounds (Plantenga and Remery 2013; Blau and Currie 2006). In a system of half-day schools, up to 38 per cent of the total weekly learning time takes place during the afternoon hours (OECD 2011). If the home environment and parental background of students are important determinants of the quality of these out-of-school-time learning activities, the differences in afternoon activities may become important drivers of inequality in learning opportunities. Therefore, a major benefit of all-day programs is the potential to enhance equality of opportunity by regulating the quantity and quality of afternoon learning activities for all children (Kuger 2006).

This line of argument led many governments to introduce more school-based afternoon activities in recent decades. In Germany, for example, all-day school programs were established and subsidized after the so-called PISA shock in 2001. Since German students performed relatively poorly in this first international student assessment of the OECD – both in terms of average achievement and equality of opportunity – politicians and experts looked for suitable reforms that could improve the school system. One of the measures adopted was the introduction of all-day schools (Kultusministerkonferenz 2002). Politicians acted upon the assumption that all-day school programs would provide particularly favorable conditions for supporting students' cognitive skills (Linberg et al. 2018).

In this paper, we investigate the extension of all-day school programs in Germany which was triggered by the Federal Government's large investment program. We analyze the effect of all-day school programs on primary school students' achievement in terms of their standardized test scores in math and German, as well as their grades and their probability to attend the academic track after primary school. We also investigate the impact on children's time use, their school satisfaction and their experience of being bullied in the classroom. In addition to evaluating the average effect of all-day schools, we conduct a heterogeneity analysis with respect to parental education, immigrant background, single parenthood, as well as gender.

To overcome the problem of non-random selection into all-day programs, we apply an instrumental variable approach that uses the exogenous nature of a funding program by the German Federal Government. Notably, the funds (four billion euros in total) could be used for all-day school related constructional purposes only.[1] The program

[1] Since the different German states are in charge of school affairs (e.g., hiring teachers), the federal

was quite successful given that the share of all-day school students has been rising dramatically ever since. In 2002, one per cent of primary school children attended an all-day program (Dehos and Paul 2021), compared to 19 per cent among the children of our study who entered primary school in 2012.

Our analysis uses detailed administrative data of the federal program containing the amount of federal money each German municipality received for their all-day school expansion. We link these data to the second cohort of the National Education Panel Study (NEPS), which provides us with longitudinal data on 5'381 primary school students, including our outcomes of interest. Funding received by the federal funding program serves as instrument for all-day school attendance of students contained in the NEPS. First stage results show that the instrument is strongly associated with all-day school attendance. We also provide evidence that the instrument is not correlated with other features of the school district which could influence student achievement.

Our second stage results reveal the following: All-day school programs positively impact upon children's achievements as assessed by their school grades. When evaluating the effect of all-day attendance on further outcomes, we find that time spent on reading increases whereas time spent on watching TV and playing computer games decreases. We also find a positive impact on children's satisfaction with their school and a negative impact on the probability of being bullied by classmates.

The heterogeneity analysis reveals that non-native speakers, children of low-educated parents as well as children with single parents do not benefit as much from all-day programs as their peers do. We thus cannot provide proof that the all-day programs significantly contribute towards decreasing inequality in the school system.

Our study relates to a growing literature investigating the impact of teaching time on student outcomes. These studies produce mixed findings. While Cannon et al. 2006 and Meyer and Klaveren 2013 find no or only temporary gains, several studies find positive effects of teaching time on student achievement (Mandel et al. 2019; Lavy 2015; Dobbie and Fryer 2013). Yet, importantly, those are most pronounced in settings with a high-quality learning environment (Rivkin and Schiman 2015). In terms of educational inequality, Huebener et al. 2017 find that the performance gaps between students widens due to increased teaching time. Similarly, also Cannon et al. 2006 finds that full-day pre-school has no additional benefit for students in families with income below the poverty threshold. However, other studies show that educational inequality could be decreased by targeting additional learning input at low achieving students (Cortes et al. 2015; Lavy and Schlosser 2005) and non-native speakers (Cannon et al. 2011). We supplement the debate by assessing whether all-day schools raise student achievement and also enhance equality of opportunity in the education system (Bellei

government is restricted in its allocation of funds to the educational system.

2009; Linberg et al. 2018; Steinmann et al. 2018) by providing causal estimates of the effects of voluntary all-day school programs.

The remainder of this article is structured as follows: Section 2 describes the primary school system in Germany as well as the federal investment program for the expansion of all-day schools. Section 3 describes the data and provides descriptive statistics. Section 4 presents the empirical framework. Section 5 presents our main findings as well as results from a sensitivity analysis. Section 6 shows the results of our heterogeneity analysis and Section 7 concludes.

2 Institutional Background

2.1 The Primary School System in Germany

School Types and Tracking in the German School System

Germany's school system is decentralized, which means that each of the country's 16 states is responsible for education. Although there are some differences across states, the general structure is uniform. The school system divides children at age ten, after four years of primary school, into three secondary school tracks: basic track (five years), middle track (six years), and high track (eight or nine years). The latter leads to the university entrance qualification and is also called academic track.

Primary schools, the school type under study, are thus especially important for the students' future educational paths since children are tracked based on their performance in fourth grade. Fourth grade teachers will recommend the highest track if they assess a child's abilities as suited to the intellectual requirements of the academic track.[2] As we use the school track attended after primary school as one of our outcomes, we shed light on the potential of all-day schools to change students' educational opportunities.

School Day Length in German Primary Schools

Traditionally, the typical primary school day has been relatively short in Germany. From grade 1 to grade 4, children's teaching time starts at 8 am and finishes around noon, so that on average children spend 4.5 hours in a half-day school on a typical day (Dehos and Paul 2021). During these hours, children are supervised, but for the rest of the day families need to organize supervision on their own. For most children in half-day schools, mothers take up the role of the main care providers in the afternoon hours

[2] In some German states, this recommendation by teachers acts as a top limit to the schooling available to the child. In the end, parents have the responsibility to choose the child's secondary school track from the (limited) set of available school tracks.
Two states conduct the tracking after sixth grade.
In our analysis we include state fixed effects, i.e., any differences regarding school track recommendations across states are controlled for.

(Felfe and Zierow 2014). Aassve et al. 2012 point out that fewer children are cared for by grandparents in Germany than in most other European countries. Also private care (e.g. nannies) traditionally plays only a very little role in Germany (Spieß et al. 2002). Moreover, publicly organized afternoon care provided by churches or the municipality (*Horte*) are only attended by a minority of children (Felfe and Zierow 2014; Dehos and Paul 2021).

After the PISA shock in 2001, the importance of introducing all-day schools across Germany attracted more attention both in the political debate as well as in the media – since all-day schools were regarded as a vehicle to increase student achievement and equality of opportunity. Yet, there has also been strong opposition to the concept of mandatory all-day school programs by the conservative party and their supporters.[3] In the end, the local authorities and the schools had to decide whether to introduce the all-day program on a mandatory or voluntary basis. Most schools opted for the latter, i.e., they did not extend the compulsory instruction time, but instead established a voluntary afternoon program. Importantly, both types of schools (with a voluntary or respectively a mandatory afternoon program) are considered as all-day school (*Ganztagsschule*) in Germany.[4]

All-day School Programs in Germany

Schools are responsible for organizing their all-day school program which is supposed to be closely connected to the primary school syllabus of the respective state (KMK 2015). At a minimum, all-day schools have to offer a program that covers seven hours per weekday (regular lessons in the morning included), and the median time of the daily program is about 8.5 hours a day (StEG 2013). This means that a child enrolled in the median all-day school would stay in school until 4.30 pm.[5]

Only about ten per cent of all-day schools follow a mandatory approach that makes school attendance in the afternoon compulsory for all children.

In schools with a voluntary all-day school program (about 90 per cent in our sample, and 86 per cent on the German average), all children attend regular lessons in the morning until noon. Afterwards, only all-day school program participants stay in school

[3]For an overview of the debate see Kuhlmann and Tillmann 2009.

[4]In our analysis we do not differentiate between mandatory and voluntary all-day programs in all-day schools. Note that the large majority of all-day schools in our sample (90%) had voluntary all-day programs.

[5]As mothers are the primary care providers in a system of half-day schools, the extended time of supervision opens up the possibility to work longer hours. Yet, in their paper on the effects of all-day schools on mothers, Dehos and Paul 2021 do not find a significant effect on maternal employment. They conclude that a lack of child care in the afternoon might not be the decisive reason why mothers do not return to full-time work once the child is in school. Additionally, mothers who are not employed when their child gets enrolled in school are rather detached from the labor market so that all-day school programs do not affect their employment status.

for lunch and activities such as sports and arts courses, recreational activities, and homework assistance. Enrollment takes place at the beginning of each school year. Parents have to pay for the child's lunch and often also have to contribute a small participation fee.

In order to provide some intuition on how the afternoon hours look like in the all-day schools of our sample, Table 1 shows statistics of the pedagogical activities and the qualification of staff employed in the afternoon hours, and contrasts these numbers with activities and staff qualification in half-day schools (all information based on questionnaires with primary-school principals in 2014, as part of the NEPS).

The table illustrates that traditional half-day schools also provide some activities in the afternoon hours, but to a much lesser extent than all-day schools. 94 per cent of all-day schools provide lunch to their students. This is the case in only 59 per cent of half-day schools. Similarly, homework supervision, sport activities, and other activities are much more common at all-day schools. When it comes to supporting learning activities for specific groups of students, 83 per cent of all all-day schools provide support for low-performing students, while about half of the all-day schools have specific support for high-performing students and non-native speakers. More additional teaching also takes place in all-day schools. For example, there are additional classes in German in 51 per cent of all-day schools, whereas this is only the case in 28 per cent of half-day schools.

In terms of employed staff in the afternoon hours, there is a large fraction of staff with no education-related qualification (58 per cent), also in all-day schools.[6]

Finally, the bottom panel of Table 1 shows combined categories calculated from the variables presented above. 52 per cent of all-day schools offer support for students at risk meaning that they have afternoon programs for low performing students and non-native speakers (only 43 per cent of half-day schools do so). 25 per cent of all-day schools offer additional classes in all of the three main subjects of primary schools, German, math and science (only 8 per cent of half-day schools do so). Finally, 43 per cent of all-day schools – in contrast to 28 per cent of half-day schools – employ an above-average share of qualified staff for afternoon activities (the sample mean share of qualified staff being 40 per cent).

Taken together, children in all-day schools are more likely to have additional classes and especially low performing students as well as non-native speakers are more likely to receive support. Yet, more than half of the staff working with the children in the afternoon hours has no qualification related to teaching. Therefore, it is not clear from

[6]In the afternoon program, an educational specialist is typically supported by several forces with different educational backgrounds, such as music and dance teachers, artists from the field of performance and design, psychological specialists, exercise instructors, educationally qualified parents, volunteers, students and pupils.

these descriptives whether a positive impact of all-day schools on school performance should be expected.

Previous research on the contents of all-day school programs in Germany raises some concerns about their quality. Pfänder et al. 2018 conclude that all-day schools have the potential to increase students' skills, but that this potential is not seized in the current system. The study of Steinmann et al. 2018 suggests that all-day schools rather provide childcare than extended formal learning opportunities. Therefore, they argue, student achievement is not supported as much as it could be, since, for example the lack of highly qualified staff limits the effectiveness of all-day schools in Germany. Sauerwein 2019 is a bit more optimistic stressing the possibilities for students' participation and integration in the German all-day school system. This can, he suggests, have a positive impact on the students' self-esteem as well as their social self-efficacy.

2.2 The Federal Investment Program and the All-Day School Expansion

By providing large subsidies via the *investment program for future, education and care* (IZBB)[7] between 2003 and 2009, the Federal Government tried to incentivize the states to invest in constructing and expanding all-day schools. The aim was to guarantee a nationwide supply of all-day school programs. The total investment volume amounted to some four billion Euros. However, the subsidies were supposed to exclusively serve constructional purposes, i.e., to build new all-day schools, and to renovate old all-day schools. Data on the use of funds show that almost 90 per cent of federal money was used to either expand existing schools (82 per cent) or to build new all-day schools (seven per cent). Ten per cent of funds were used for other one-time investments, e.g., for quality development measures. The federal funds were transferred to more than 8'200 schools that applied for funding with a concept on how they would use the money. More than two billion Euros were allocated to primary schools (BMBF 2009). The program was successful in raising the share of primary school students attending an all-day school. Their share increased from one per cent to 22 per cent (19 per cent in our sample) within ten years following the implementation of the federal investment program (Dehos and Paul 2021).

Yet, the investment allocation was not uniform across municipalities and states. In line with that, also the expansion of all-day slots varies greatly across regions. One possible reason could be that – in contrast to the one-time federal investments – all running costs of all-day programs must be paid by the states and municipalities. They had to cover the personnel costs, e.g., teacher salaries, and operating costs, e.g., expenses for

[7] *Investitionsprogramm Zukunft Bildung und Betreuung*, for more information see https://www.ganztagsschulen.org/de/868.php

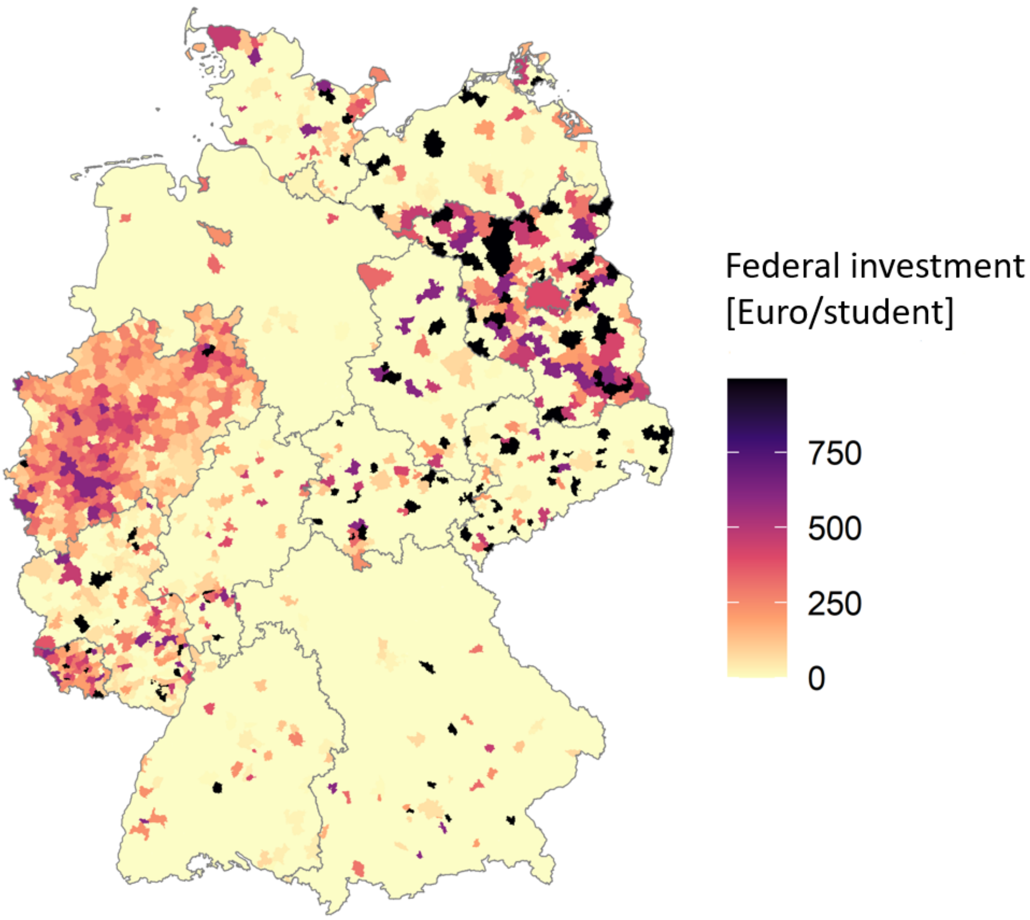
Table 1: All-day schools versus half-day schools: pedagogical content and staff qualification of employees in the afternoon hours

	half-day	all-day	difference	t-statistic
percentage in sample	.463	.537		
lunch	.59	.943	.352	34.291
activities offered at least once per week:				
home work supervision	.606	.93	.324	31.022
sport activities	.599	.821	.222	18.61
artsitsic activities	.679	.82	.14	12.084
mechanical activities	.35	.591	.241	18.2
computing activities	.38	.554	.174	12.981
support for:				
high perfoming students	.338	.496	.158	11.855
low performing students	.645	.825	.181	15.427
non-native speakers	.458	.548	.09	6.642
additional classes in:				
maths	.216	.341	.125	10.259
science	.304	.459	.155	11.822
German	.284	.511	.227	17.344
foreign languages	.228	.284	.056	4.687
employees in afternoon programs:				
no related qualification	.571	.575	.004	.526
qualified in childcare	.249	.288	.039	5.061
qualified in teaching	.181	.137	-.043	-7.93
share of students attending school with:				
support for student at risk	.433	.523	.091	6.669
additional classes in main subjects	.076	.246	.169	17.051
a high share of qualified staff	.279	.426	.147	11.372
N	2494	2887		

The mean shares for the groups of employees are calculated from all non-missing observations. All other variables are equal to one if the principal reports to offer this activity and zero if not. The lower panel aggregates parts of the information above: Supporting students at risk is one if the school has support for low performing students and non-native speaekrs. Offering additional classes in main subjects is one if the school offers classes in math, science and German. Having a high share of qualified staff refers to a share of employees in afternoon programs which is qualified in teaching or childcare above the sample mean.
Source: NEPS SC2, own calculations based on answers from the principal's questionnaire of wave 4 conducted in 2014.

lunch, of the all-day school expansion. The personnel costs for one additional all-day primary school student are estimated between 992 Euro and 1'981 Euro per year, depending on the intensity of the all-day program (Klemm and Zorn 2017). Therefore, we control for pre-expansion municipality characteristics that are related to the economic means of a municipality (fiscal capacity) and the demographic need for all-day schools (number of inhabitants, female labour market participation). Other regional variables do not predict whether a municipality received funding (see Table A.1 in the Appendix). A potential reason for remaining regional variation is the application procedure which has sometimes been facing administrative restrictions. Exogenous factors like dates of the yearly school conference, or complicated communication procedures between the different government agencies could have slowed down or even prevented the reception of funding.[8] Figure 1 shows the resulting regional variation of received funds from the federal investment program. We use this variation for our empirical strategy.

Figure 1: Federal investment per student across municipalities



Federal all-day school investments per student calculated with data of SPI NRW (2010)

[8]In order to receive the all-day school funding, local authorities first had to apply to their respective state which was responsible for the selection of appropriate all-day programs. The school concept had to be developed by the school director and had to be approved by the school council. In addition, in some states parental consent was also needed so that the concept had to be approved by the parent council at the school conference (see Nemitz 2016 for a comprehensive summary of the application procedure).

3 Data and Descriptive Statistics

3.1 Data

We rely on three data sources to evaluate the impact of all-day schools on student achievement. The first data source is the second cohort of the National Education Panel Study (NEPS). It provides us with data on the achievement and background characteristics of primary school students in Germany. The second data source is the administrative register of the federal all-day school investment program (IZBB). The third source is the INKAR database containing administrative data on municipality characteristics. We merge all three data bases at the municipality level.[9]

National Education Panel Study (NEPS)

The second cohort of the NEPS is an annual panel collecting rich data on the educational progress of young children in Germany. More than 6'000 participants in over 300 schools are sampled at primary school entrance in 2012. Our study includes five waves of this cohort covering their four years in primary school and their first year in high school.[10] We exclude cases with missing values in all outcome variables or with missing municipality of residence as we are unable to match the instrument without the latter. This leaves us with 5'381 observations.

We define two alternative treatment variables: first, we use the individual all-day program attendance as treatment variable. Second, as alternative definition, we use the share of students in a respective school who attend an all-day program, i.e., the *all-day share*, as treatment variable.[11] Defining the treatment variable on school level – rather than on individual level – has the advantage that it accounts for the possibility of spillovers within schools. Students who do not attend the afternoon program could be influenced by the existence of that program if a sufficiently large number of classmates does. These estimates incorporating potential peer effects complement the estimates on individual level.

Each wave consists of surveys with children's parents, their class teachers as well as school principals, and from third grade onwards also with the children themselves.

[9] For the merge with the NEPS data, we must categorize the instrumental variable and the municipality characteristics and are not allowed to use the continuous value. This is in order to align with the NEPS data protection rules and aims at preventing us from identifying any municipality through our matched information.

[10] For a sub-sample of observations, information on their time in kindergarten would be potentially available as well. However, conditioning on the availability of information from kindergarten would add selectivity and reduce the sample size below 1'500. Therefore, we decided to limit the analysis to the primary school and high school outcomes.

[11] As the expansion of all-day schools was gradual and the attendance to all-day programs is not mandatory, only a few schools have 100 per cent of students enrolled in an all-day program.

Additionally, for each wave the NEPS researchers designed standardized tests in various school subjects taken by the surveyed children (Blossfeld et al. 2011). These standardized tests are an important element of our study and a major benefit of the NEPS data, as the NEPS is the first (and only) data source providing nation-wide panel data on student achievement in Germany.

In fourth grade, the students are tested in German and in math. We use these standardized test scores at the end of primary school as outcome variables in our main regressions.

Besides the achievement tests, we use the final grades students received in German and math in fourth grade to measure educational success. These subjective assessments of students' achievement may capture non-cognitive skills known to be important predictors for outcomes later in life (Jackson 2018; Brookhart et al. 2016; Bowers 2011), but not captured by standardized tests. We therefore include the grades as outcomes in our main analysis.

As described in Section 2.1, the performance in fourth grade is relevant for the tracking of students after primary school. We therefore use the attendance of the high track in fifth grade (*Gymnasium*, leading to the university entrance exam) as further outcome variable to assess the impact of all-day schools on the educational trajectory.

By making use of further information provided in earlier waves of the panel, we also look at the following outcomes in an additional analysis (see Section 5.5): We evaluate whether attending an all-day school influences enrollment in a sports club and time spent on homework after school (with data from wave 4 when students are in second grade). Furthermore, we analyze whether attending an all-day school influences students' satisfaction with their school, the probability that they are being bullied, and their propensity to do regularly physical exercise (with data from wave 5 when students are in third grade). To assess further effects on time use and substitution of activities, we use the assessment of parents on how much time they spent with their child (when children are in third grade), and we estimate whether attending an all-day school affects time spent on reading and time spent on watching TV (when children are in fourth grade).

From the different questionnaires, we gain further valuable control variables at the individual, teacher and school level. At the individual level, we control for being female, a non-native speaker, having parents with an academic degree, living in a single parent household, and having siblings. We further include the student's age in months. At the teacher level, we include dummies indicating the teacher to be male and to work full-time. As proxy for their teaching experience, we further use the teacher's age. At the school level, we build control variables for being a private school, the share of students with a migrant background, and for having more students from low socio-economic status (SES) than from high SES.[12]

Investment Program Data

Data on the federal investment program for the construction of all-day schools (IZBB) were collected and made available by SPI NRW 2010. They contain detailed information on the amount and year of spending for each school that received funding. We aggregate the investments over time and by municipality. We thus build a variable for the total investment from 2003 to 2009 per municipality.

In order to merge the investment data with the student achievement data, we have to minimize their informational content due to data protection rules. For this purpose, we first calculate the relative investments per student on municipality-level and form the distribution of relative investment of the municipalities weighted by the number of inhabitants. We then divide the part of the distribution with positive investments in 20 units of five per cent (vigintiles). For each of these vigintile categories, we calculate the median. The resulting median of vigintile categories serves as our instrumental variable.

Figure 2 displays the distribution of the instrumental variable. Considering all 4'500 municipalities in Germany, about one half of them received positive funding, see left panel of Figure 2. The right panel of Figure 2 shows the distribution of received funding of the NEPS-observations who live in more than 400 municipalities from all German states. Compared to the full population, the overall pattern is highly similar.

Municipality Characteristics

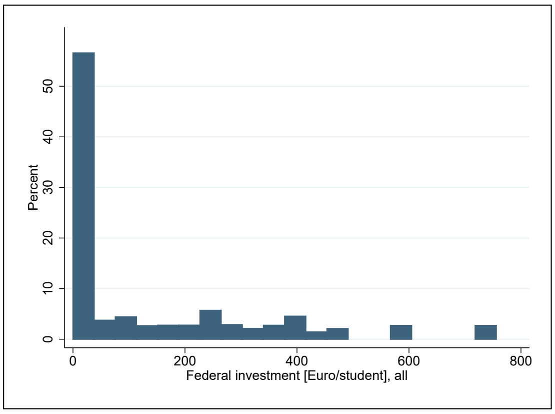
Furthermore, we use municipality level information on female labor market participation, the number of inhabitants and tax capacity. For all of these controls, we use data from 2003, i.e., before the investments took place. Again, for merging these data with the student achievement data, we have to minimize their informational content due to data protection rules. Therefore, we transform them into categorical variables. We round the female labor market participation rate to even percentage shares (resulting in 15 distinct values). We round the number of inhabitants to multiples of 80'000 and build three catgeories for municipalities below 40'000 inhabitants (in total ten distinct values) and we round the the tax capacity to multiples of 300 Euro per capita (five distinct values).

Note that we do not employ municipality fixed effects in our analysis because the federal investment is received on the municipality level and the variable does not vary

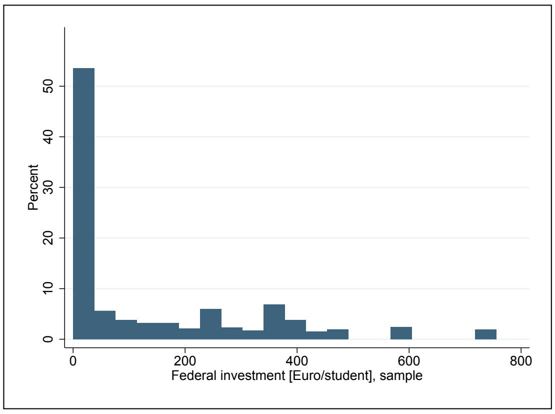
[12] In their questionnaire, the school principals are asked to provide an estimate on the two categories of demographic composition.

over time. Furthermore, counties, the next largest administrative unit, include only information on one school in more than 50 percent of cases for our data, meaning that county and municipality information merged to the NEPS data uniquely applies to one school and their students in most cases, which also prevents us from employing county fixed effects. We opted therefore for state fixed-effects.

Figure 2: The distribution of the instrumental variable



(a) all municipalities



(b) sample

Source: SPI NRW 2010; BBSR 2018; Destatis 2017

Left: Relative investments per student on municipality level as median of vigintile categories for all German municipalities (weighted by inhabitants).

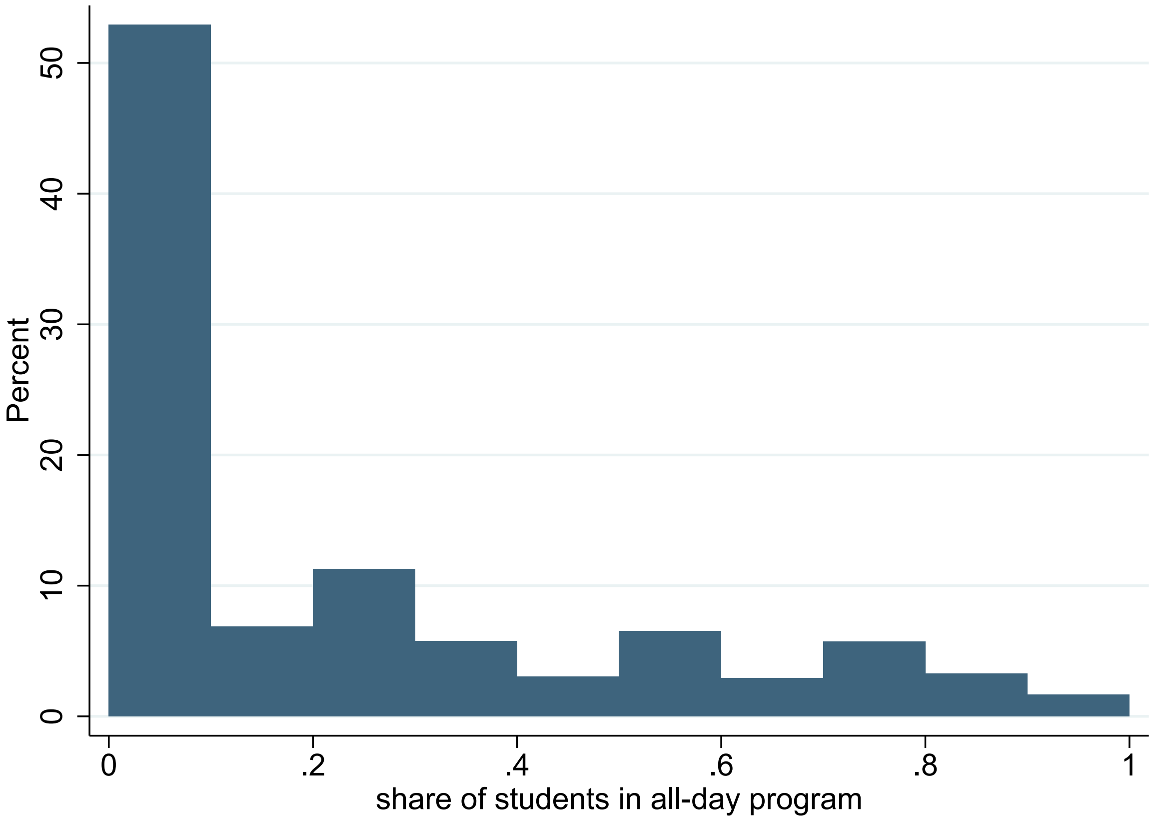
Right: Relative investments per student on municipality level as median of vigintile categories for the observations in the sample.

3.2 Descriptive Statistics

Figure 3 illustrates the variation of the treatment by displaying the distribution of children over the share of all-day slots in their schools. About half of observed children attend a school with an all-day share of zero, meaning that 100 per cent children in their schools only attend the half-day program. The other half of observed children attend a school with a non-zero all-day share, with most of them having a share below 60 per cent, meaning that 60 per cent or less children in their school attend the all-day program. Only few children attend schools with 100 percent students enrolled in the all-day program.

Table 2 shows the descriptive statistics of individual and family characteristics, school characteristics, teacher characteristics, municipality characteristics, as well as the outcome variables. The first column shows averages for the whole sample, columns 2 and 3 show the subgroup means for the group of children attending an all-day school and those attending a half-day school (note that for this table we define all-day schools

Figure 3: Share of all-day slots in observed schools



Source: NEPS SC 2.
Share of students attending an allday-program on school level.
Example: In roughly 50 percent of cases 0 to 10 percent of the students in the school attend an allday program.

as schools having a non-zero all-day share). The right-hand part of Table 2, columns 5 to 8, indicate the differences between the different subgroups of students. They demonstrate the differences in means between girls and boys, non-native and native speakers, children whose parents do and do not hold a university degree, as well as children raised by a single parent and those raised by two parents.

For the individual characteristics, we find that students in all-day schools more often live in households with only one parent and without siblings. As regards school characteristics, the share of students with a migrant background is significantly higher in all-day schools. Furthermore, teachers in all-day schools are more often male, older, and more often work full-time. Additionally, we see that all-day schools are more common in larger municipalities (in terms of inhabitants), and with a higher tax capacity.

Regarding outcome variables, there are no significant raw differences between the students of all-day and half-day schools. The differences between subgroups show some strong but not very surprising patterns. While girls do worse in math than boys, they outperform them in language skills. Children of parents with a university degree show generally above average results, and children raised by a single parent perform below the average. Non-native speakers perform worse than native speakers in standardized tests and in school grades.

Table 2: Descriptive statistics for control and outcome variables

	entire sample	all-day	half-day	t-statistic	differences between subgroups female	non-German speaker	university diploma, parents	single parent
individual controls								
female	.515	.503	.528	-1.118		.018	-.013	.007
non German speaker	.172	.178	.166	1.398	.01		$-.065^{***}$	$.031^{*}$
university diploma, parents	.58	.567	.596	-.869	-.013	$-.11^{***}$		$-.244^{***}$
single parent	.111	.119	.102	2.049	.003	$.022^{*}$	$-.1^{***}$	
age at school entry (months)	77.299	77.061	77.574	.58	$-.789^{***}$	-.158	$-.916^{***}$	$.612^{***}$
living with siblings	.709	.701	.719	-2.586	.002	$-.049^{***}$	$.124^{***}$	$-.271^{***}$
school controls								
private school	.042	.043	.041	1.063	.008	$-.015^{**}$	$.017^{***}$	-.003
low SES school	.337	.345	.327	.229	-.015	.011	$-.08^{***}$	.032
migrant share	.233	.271	.189	10.632	.001	$.028^{***}$	$-.047^{***}$	$.016^{**}$
teacher controls								
male teacher	.064	.069	.057	2.566	-.003	$-.017^{*}$	$-.011^{*}$	-.006
teacher age	46.817	47.615	45.894	3.547	-.013	$-.537^{*}$	.29	-.103
full-time teacher	.292	.306	.275	1.74	-.005	.017	-.019	-.003
municipality controls								
female labor market partipation	44.269	43.57	45.08	-.302	-.039	.098	.113	-.127
inhabitans rounded, in 100'000	1.215	1.568	.806	7.469	-.037	$.224^{***}$	$.075^{*}$	.087
taxable capicity munc.	4.78	4.86	4.687	4.193	.019	$.203^{***}$	$.172^{***}$	-.044
Outcomes								
math test scores	.062	.039	.089	-1.878	$-.119^{***}$	$-.14^{***}$	$.5^{***}$	$-.238^{***}$
German test scores	.052	.006	.106	-1.177	$.221^{***}$	$-.087^{**}$	$.472^{***}$	$-.212^{***}$
School grade math	4.882	4.86	4.907	-1.346	$-.061^{**}$	$-.062^{*}$	$.339^{***}$	$-.275^{***}$
School grade German	4.841	4.823	4.862	-.646	$.272^{***}$	$-.056^{*}$	$.316^{***}$	$-.198^{***}$
high track	.599	.593	.605	-1.151	.012	.015	$.231^{***}$	$-.156^{***}$
N	5381	2887	2494					

Source: NEPS SC2, own calculations
The individual control variables are dummies for being female, having another mother tongue than German, for living with at least one parent holding a university degree, in a single-parent household, with siblings in the same household and a variable examining for age at school entry in months.
The school level control variables are dummies indicating whether it is a private school and whether the school has more students from low SES than from high SES and the share of students from migrant background (coded as median if missing). The latter numbers are given by the principal.
The teacher controls are dummy variables indicating whether the teacher is male and working full-time and the teacher's age (coded as median if missing).
The municipality controls are categorical variables for the female labor market participation (rounded to even percentage shares, 15 categories), number of inhabitants (rounded to 80'000 inhabitants and with three categories for those municipalities which would be rounded to zero, ten categories) and the fiscal capacity (rounded to 300 Euro per capita, five categories). All four municipality variables give the numbers of the year 2003, before the intervention started.
The outcomes are standardized test scores for math and German (constructed as mean of orthography and reading test scores), school grades from 1 (fail) to 6 (very good) and a dummy for attending the high track of high school.
Column 1 shows the distribution of the controls for the entire sample. The middle panel is divided in subgroups according to availability of all-day programs in schools (e.g., having a positive all-day share). The t-statistic belongs to the test whether the difference is significant in a regression with state fixed effects.
The right-hand panel gives the variation according to subgroups of students. The columns show the difference in means between the group having value one and zero for the respective subgroup identifier. The stars indicate significance on ten, five and one percent level in a regression controlling for state fixed effects.

For a better understanding of how different subgroups of students are represented in all-day schools, Table 3 provides some details on the number (columns 2 and 3) and shares of students in the different school types (columns 4-6). The fourth column demonstrates that the probability to attend an all-day school is very similar across all subgroups (between 52 and 57 per cent). Since not all students enrolled in an all-day school attend an all-day program, due to the latter's voluntary nature or supply constraints, the shares of subgroups look different when investigating the participation in an all-day program conditional on attending an all-day school (Column (6) of Table 3). The probability of attending an all-day program once enrolled in an all-day school is 46 per cent for a child with a single parent and about ten percentage points lower for the other subgroups. As the children from single-parent households have the highest share in all-day schools and the highest "conditional share", they have also the highest share of students in all-day programs with 27 percent. That could be explained by supply as well demand side behavior. In the case of oversubscription to all-day porgrams, school administrations could allocate preferentially children from single parents to all-day slots and single-parents may feel higher needs to prolonged schooling/care for their children.

Table 3: Numbers and shares of individual attendance of all-day schools and all-day programs

	number of observations			shares		
	total	in allday-school	in allday-program	in allday-school	in allday-program	conditional share
subgroups:						
female	2769	1453	517	.52	.19	.36
non German speaker	927	513	193	.55	.21	.38
university diploma, parents	3123	1637	597	.52	.19	.36
single parent	599	344	159	.57	.27	.46
living with siblings	3815	2023	723	.53	.19	.36
entire sample	5381	2887	1028	.54	.19	.36

Source: NEPS SC2, own calculations
The left hand panel shows number of observations in all-day schools and all-day programs for the different subgroups.
The right hand panel shows the share of observations in all-day schools and all-day programs for those subgroups. The conditional share indicates the share of children in all-day programs conditional on attending an all-day school.

4 Empirical Strategy

Our analysis aims to estimate the causal impact of attending an all-day school on student achievement. The general challenge of analyzing the effects of all-day programs is that the enrollment into these programs is endogenous. We address this challenge by using an instrumental variable strategy. For our instrument, we use data on the federal all-day school investment project that provided funding for constructional purposes (see Section 2.2).[13]

It is naturally to think that the federal investments were not distributed randomly across the country. Schools and municipalities needed to apply and could therefore select into the funding program. To account for this potential self-selection, we examine in Table A.1 which municipality characteristics have a significant impact on the amount of funding. It turns out that federal funding increases significantly with the number of inhabitants and the tax capacity. Further, the female employment rate has a positive

[13]Reassuringly, the same investment program has been used as an instrumental variable for all-day school attendance by Dehos and Paul 2021 and Nemitz 2016 in their studies on the impact of all-day schools on female employment.

effect which is at least marginally significant. These findings are not surprising. On one hand, larger municipalities and those with higher tax revenue have better capabilities to bear the subsequent costs for running the all-day schools. On the other hand, for municipalities with a higher employment rate of women, there is greater need for childcare in the afternoon and therefore local authorities may have a higher propensity to apply for funding. Nevertheless, our regression shows that there are many variables which could be intuitively important but do not seem to have a large influence, among them the GDP p.c., the ratio of apprenticeships to applicants (both on county level) and birth rate in the municipality. We tested 11 further variables in total which are all insignificant, individually and jointly. Consequently, we included in our approach only the number of inhabitants, tax capacity and the female employment rate (see A.1).

For our project, we use the panel data of the NEPS in a cross-sectional manner. The all-day programs can be thought of as an intervention which happens throughout primary school. Therefore, we use control variables from the begin and the outcomes of the end of primary school (grade 4). Since most children attend all-day programs during their entire time at primary school or not at all, this approach is suitable, and we hence use the sample as a cross-section.[14]

The all-day school attendance of student i in school s in municipality m – $Allday_{ism}$ – is regressed on the instrumental variable $Investment_m$ (received investment funds in municipality m), control variables for school (s) and municipality (m) characteristics as well as students' characteristics i in school s – X_{ism}. In addition, we include state fixed effects δ_b. The estimated coefficient α is hence the effect of an additional unit of investment on the probability to attend an all-day school. For reasons of simplicity, we assume that this effect is linear. When we use the alternative treatment, all-day share on school level, as robustness check, the estimated coefficient shows the effect of an additional unit of investment on the share of all-day students in a given school. For defining a robust all-day share, we discarded the schools from the bottom vigintile in terms of the number of valid information on all-day attendance.

Our first stage estimation looks as follows:

$$Allday_{ism} = \pi' X_{ism} + \alpha Investment_m + \delta_b + \eta_{ism} \tag{1}$$

In our setting, the exclusion restriction requires that the instrumental variable, received investment funds, only impacts student outcomes through the channel of expanded all-day schooling. We argue that this requirement is fulfilled because (1) funding was only allowed to be used for constructional expenses and not for spending on

[14]Whether there are children who join or drop out of all-day programs from one grade to another is not observable in our data. However, using the changes in the all-day status in a panel setting would not solve the endogeneity concerns and they are unlikely to happen at a high frequency.

teachers' salaries and other important factors of the education production function, (2) school catchment areas are binding for primary students in Germany, i.e. parents would have to move in order to enroll their child into a school with funding if their current school catchment area had not received funding, and this type of mobility is very low in Germany.

As support for the results of the first stage, we report Oster bounds (Oster 2019). They give a measure of how much unobserved confounders would bias our estimate. Of course, we cannot exclude that factors we are not able to control for have explanatory power on the likelihood to attend an all-day program. Despite our control regression which shows that many different regional variables are unrelated to the instrumental variable (see table A.1), it is conceivable that unobservables have an influence on the all-day attendance. However, the Oster bounds indicate that additional variables even if they would explain as much of the outcome as the included controls, would not largely bias the estimate on the investments.

After the first stage we show reduced form results, the regression of the educational outcomes on the investments. The influence of investments on the performance of students is highly relevant. It relates directly the public investment to the school outcomes.

Finally, using the instrumented all-day attendance, our second stage looks as follows:

$$y_{ism} = \beta' X_{ism} + \gamma \widehat{Allday}_{ism} + \delta_b + \epsilon_{ism} \tag{2}$$

Thus, the student achievement outcome y_{ism} of student i attending school s in municipality m is regressed on the instrumented all-day school attendance $Allday_{ism}$, while controlling for individual, school, and municipality characteristics X_{ism} as well as state fixed-effects δ_b. The coefficient γ shows the effect of all-day school attendance – following a LATE interpretation, the effect of the all-day school attendance that is due to the federal investment program – on student i's educational outcomes y. We contrast the estimates on the individual all-day attendance with the finding on the all-day share on school level.

5 Results

In this section, we present the results of our IV estimations. We begin with the first stage estimating the effect of the instrument (federal investment for all-day schools) on the endogenous variable (all-day school attendance). Then, we show the reduced form results estimating the effect of our instrument on the outcomes of interest. Finally, we show the second stage results, the causal effect of the instrumented endogenous variable on the outcomes of interest. The section continues with a sensitivity analysis and an analysis of the effect of all-day school attendance on further outcomes.

5.1 First Stage Results

Table 4 shows the estimates for the first-stage regression (see Equation 1). In this table, we compare the first stage results with and without controls, which provides some insights into the relevance of selection on observables with respect to the instrument.

The main variables of interest, individual all-day program attendance (columns 1 and 2), as well as the share of all-day students in a school (columns 3 and 4), are regressed on the federal investments per student (and additionally on control variables in column 2 and 4, respectively). Most importantly, we find that the instrument is strongly significant and economically meaningful. An increase in the investment of 1'000 Euros per student is estimated to increase the probability to attend an all-day school by about three percentage points. Moreover, we do not encounter a weak instrument problem with an F-statistic larger than 17 in the regressions including the control variables.

The bottom of Table 4 reports Oster bounds (Oster 2019). The Oster bounds define a limit to which unobservables could bias the coefficient for the variable of interest. The parameter delta is set equal to 1 which means that the unobservables are allowed to explain the outcome as strongly as the observables. We further admit these unobservables to improve the explanatory power of the model measured by R2 by 30 percent (such that the R2 of the hypothetical model is 1.3 times the R2 of the regression with the available control variables). Based on this, we can evaluate by how much the coefficient beta of our variable of interest (the federal investments) can change under these two assumptions. It is also important according to Oster's definition that the bias of the added (un-)observables towards an uncontrolled regression keeps the same sign and that the estimated bound does not differ more than 2.8 times the standard error from the original beta. Both conditions are given in our case for both first stages – using individual all-day participation as outcome (column 2) as well as when using the all-day share as outcome (column 4). In both cases, the Oster bounds are still distinctly positive which adds confidence in the power of the instrument.

When it comes to control variables, we see some significant coefficients even though the inclusion of control variables does not change the size or significance of the coefficient of the federal investment variable by much (comparing columns (1) and (2), and columns (3) and (4) respectively). The results indicate that non-native speakers have a higher likelihood of attending an all-day program. Students without siblings and students from a single-parent household also attend all-day schools more often. The probability to attend an all-day school is also higher for a student living in a municipality with more inhabitants and with a higher tax capacity.

Table 4: First Stage: Regression on individual all-day attendance and the share of all-day students in school

	individual all-day attendance		share of students in all-day program	
	(1)	(2)	(3)	(4)
investment	.4176***	.3328***	.3962***	.3178***
	(0.000)	(0.000)	(0.000)	(0.000)
female		.001907		-.003868
		(0.869)		(0.445)
non-native speaker		.03141**		.0081
		(0.040)		(0.359)
age in months		-.001493		.000266
		(0.265)		(0.689)
parent, college degree		-.005703		.0129
		(0.663)		(0.112)
single parent household		.101***		.0226**
		(0.000)		(0.024)
sibling in household		-.1148***		-.01939**
		(0.000)		(0.013)
private school		.11		.1177
		(0.422)		(0.403)
low SES school		-.03001		-.03256
		(0.333)		(0.295)
share of migrants in school		.09104		.1269
		(0.408)		(0.252)
male teacher		.001637		-.001751
		(0.970)		(0.955)
teacher age		-.000577		.000206
		(0.581)		(0.821)
full-time teacher		.02858		.04079*
		(0.204)		(0.062)
female employment rate		-.001694		-.000084
		(0.626)		(0.979)
inhabitans (100'000s)		.02221**		.02185**
		(0.040)		(0.048)
tax capicity (100 Euro p.c.)		.01218*		.009076
		(0.070)		(0.163)
Constant	-.1331**	.05227	.2921***	.05237
	(0.037)	(0.761)	(0.000)	(0.819)
N	4502	4502	5381	5375
Fstatistic	87.6	17.06	290.3	16.81
R2	.17	.223	.375	.445
Oster bound ($\beta^*(R_{max} = 1.3\tilde{R}, \delta = 1)$)		0.209		0.144

p-values in parentheses

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

Dependent variable in columns (1) and (2) is a dummy for attending an all-day program.
Dependent variable in columns (3) and (4) is the share of students in all-day programs in the school ranking from zero to one.
Investments are given in 1000 Euro per student as median of vigintile categories of the distribution of relative investments per student.
The F-statistic refers to a test for exclusion of the instrument.
For a more detailed explanation on the control variables, see notes of table 2.
Control dummies for missing information on share of students with migrant background, SES of students and teacher's age, gender and working hours and state fixed-effects are included. Standard errors are clustered on school level.
The last column gives the "Oster bound": The estimated coefficient for investments under the assumption that there exist unobserved controls which would increase the R^2 of the full-model ($\tilde{R}$) by 30 percent and which have the same explanatory power as the already included controls (δ) (Oster 2019).

5.2 Reduced Form Results

Although we ultimately want to estimate the causal effect of all-day school attendance on educational outcomes, it is also interesting to look at the reduced form effects of the federal investment program. The reduced form effects shown in Table 5 illustrate the direct impact of received federal investment for all-day schools on students' educational outcomes. We use standardized test scores and school grades in fourth grade in math and in German as outcome variables. The fifth outcome is a dummy whether the student attends the academic track in high school.

Results in Table 5 indicate that the investment has a positive significant impact on students' grades. An increase in funding of 1'000 Euro per student improves the grade in math and German by 18 and 25 percent of grade point (on a scale with six grade points), respectively. The probability to attend the academic track after primary school is also significantly increased by 19 percentage points for 1'000 Euro per student.

In sum, the reduced form results contain evidence that the federal investments for all-day schools provided to municipalities had an overall positive impact on the respective municipalities' students although the effect on standardized test scores is insignificant.

Table 5: Reduced form regression: The educational outcomes are regressed on the investments

	test scores		school grades		
	maths	German	maths	German	high track
investments	.112	.027	.18*	.246**	.189***
	(0.376)	(0.847)	(0.095)	(0.027)	(0.004)
female	-.125***	.219***	-.0575**	.275***	.0136
	(0.000)	(0.000)	(0.030)	(0.000)	(0.399)
non-native speaker	-.0744**	-.033	-.0296	-.0277	.0309
	(0.040)	(0.317)	(0.432)	(0.423)	(0.153)
age in months	-.013***	-.00924***	-.00809**	-.00514*	-.0063***
	(0.000)	(0.005)	(0.018)	(0.087)	(0.000)
parent, college degree	.438***	.43***	.311***	.303***	.213***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
single parent household	-.079*	-.0813*	-.193***	-.133***	-.104***
	(0.052)	(0.052)	(0.000)	(0.003)	(0.000)
sibling in household	.102***	.0477*	.00691	-.0399	.0184
	(0.001)	(0.097)	(0.810)	(0.109)	(0.297)
private school	.00429	.1	-.142*	-.158**	-.0332
	(0.964)	(0.219)	(0.056)	(0.023)	(0.461)
low SES school	-.0389	-.0577	.00808	-.069*	-.0367
	(0.361)	(0.157)	(0.811)	(0.053)	(0.148)
share of migrants in school	-.341***	-.242**	-.041	-.0418	.0906
	(0.001)	(0.038)	(0.638)	(0.661)	(0.145)
male teacher	-.00767	-.0379	.0487	.0956**	.053
	(0.898)	(0.513)	(0.278)	(0.048)	(0.106)
teacher age	.00465***	.00476***	-.00178	-.00196	-.00106
	(0.005)	(0.007)	(0.274)	(0.199)	(0.329)
full-time teacher	-.00689	.0534	-.00253	.0117	-.00769
	(0.868)	(0.178)	(0.941)	(0.745)	(0.722)
female employment rate	.00537	.00446	-.000738	.00061	-.00465
	(0.334)	(0.422)	(0.876)	(0.902)	(0.140)
inhabitans (100'000s)	.00863	.00598	-.00516	.00444	-.00547
	(0.572)	(0.711)	(0.585)	(0.653)	(0.417)
tax capicity (100 Euro p.c.)	.00922	.00878	.00815	.00604	.0219***
	(0.393)	(0.391)	(0.368)	(0.488)	(0.000)
N	5087	5200	4062	4059	3933

p-values in parentheses

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

The dependent variables are standardized test scores, school grades from 1 (fail) to 6 (very good) and attending the high track in high school.

Investments are given in 1000 Euro per student as median of vigintile categories of the distribution of relative investments per student.

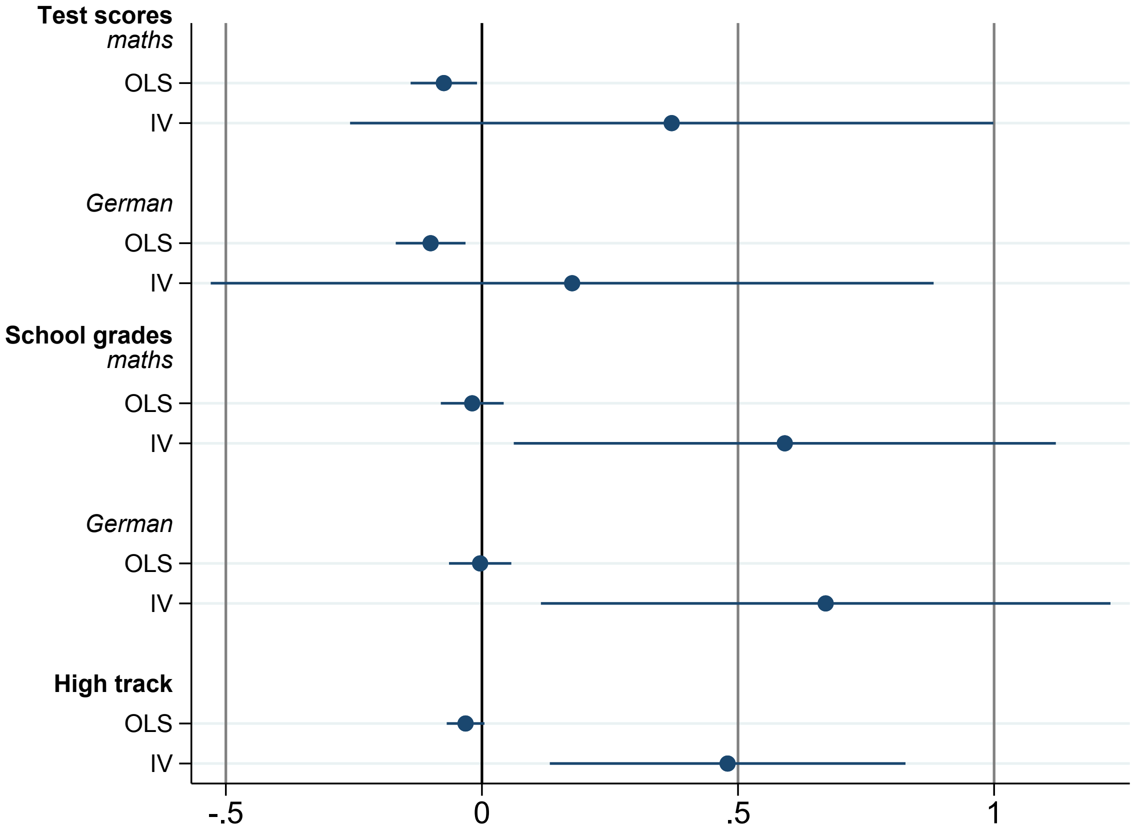
For a more detailed explanation on the control variables, see notes of table 2.

Control dummies for missing information on share of students with migrant background, SES of students and teacher's age, gender and working hours and state fixed-effects are included. Standard errors are clustered on school level.

5.3 Second Stage Results

Figure 4 shows the second stage results of our instrumental variable approach. We contrast the IV estimates with simple OLS estimates. The latter are biased if all-day school attendance is endogenous. The results suggest that this is indeed the case. All OLS estimates are negative and differ from our causal IV estimates in their sign. Selection into all-day schools thus seems to be driven by students with more disadvantages regarding their ability or other relevant unobserved characteristics that have an adverse impact on educational outcomes. This underlines that selection is an important issue when it comes to the evaluation of all-day schools.

Figure 4: The impact of individual all-day attendance on student achievement



Notes: 2SLS estimates for individual all-day attendance (zero-one dummy variable) and the 90 percent confidence intervals.
The dependent variables are normalized test scores, school grades (from 1(fail) to 6(very good)), and attendance of high track in high school.
State fixed-effects and a set of individual, teacher, school and municipality level controls and controls for missing values in those variables are included. Standard errors are clustered on school level.
See tables A.2 and A.3 for more detailed results.
Source: NEPS SC 2.

The IV results show the impact of all-day school attendance on fourth-graders' test scores in math and German, on their grades in math and German, and on being enrolled in the high track after primary school. For a more detailed regression output see Tables A.2 and A.3 in the Appendix.

While the estimated effects on students' test scores are insignificant, there are positive and significant effects of all-day attendance on students' grades as well as on the probability that students' attend the high track after primary school.

The results indicate that all-day school attendance leads to an increase of about 59 and 67 percent of a grade on students' math and German report cards (on a scale from 1 to 6). Furthermore, the estimate for the effect on high-track attendance is 0.48. This means that a student, who attends an all-day school because of the federal investment, has a 48 ppt higher probability to attend the high track in secondary school.

We thus find significant effects on the grades and high-track attendance, while the effect on test scores is insignificant. This result might seem puzzling but is not self-contradictory. Perceived achievement by the teacher measured in grades may capture the pure knowledge in a subject and additionally some non-cognitive skills which are not measured by test scores. It is therefore possible that all-day programs tend to influence those skills that are important for the teacher assessment (grading and tracking), but are not detectable in standardized tests. This is line with findings in other studies on the informational content of grades, e.g., Jackson 2018, Brookhart et al. 2016, and Bowers 2011.

In order to account for the possibility of spillovers of all-day programs within schools, we use the alternative treatment variable *all-day share* (the share of students attending the all-day program in a school). Students who do not attend the afternoon program could be influenced by the existence of that program if a sufficiently large number of classmates does. These peer effects are neglected using individual all-day school attendance as treatment variable. Figure A.1 shows that a higher all-day share – induced by the federal investment program – has a positive impact on students' outcomes. The interpretation of the coefficients is, of course, different from the previous one. For example, the estimate for the effect of the all-day share on high-track attendance is 0.51. This means that an increase in a school's all-day share by ten percentage points increases the probability for a student to attend the high-track by approximately five percentage points. Generally, the results point in a very similar direction. For the all-day share, the estimates on the tracking and the German grade are significant while the ones for the test scores are again not significant.

5.4 Sensitivity Analysis

As shown in Table 1 in Section 2, it is not only all-day schools that offer afternoon activities to their students: Some half-day schools also provide voluntary activities and student support in the afternoon. Therefore, not accounting for the existence and quality of afternoon programs in half-day schools (that are the counterfactual scenario to all-day schools) could downward bias the estimated effect of attending an all-day school. In Table A.4 in the Appendix, we include content and staff qualification of afternoon programs as control variables – in separate regressions as well as combined in one regression. In the first row, we display the results when not examining any afternoon program characteristics. In the second row, we examine the existence of additional classes in main subjects taking place in the afternoon, and we find that all coefficients of the treatment variable increase in size. This also applies when holding support for children with low achievements and non-native speaker constant (fourth row). When accounting for the share of qualified employees (third row), the estimated coefficients deviate a bit in both directions from the baseline. However, neither of the three controls leads to strong changes in any of the five point estimates or the significance levels.

When jointly accounting for all three types of afternoon program characteristics, the respective coefficients of the treatment variable increase compared to the first row (fifth row). As regards the effects on standardized test scores, the coefficients on the math score become a bit larger but do not gain much significance. In contrast, effects on high track attendance and the school grades increase when adding the quality controls and stay significant.

These results indicate that the effects of all-day schools shown in Figure 4 cannot be replicated by a system of half-day schools providing certain elements of afternoon programs. The gains seem to be exclusively related to the institution of all-day schools.

5.5 Further Outcomes

While we use the outcomes of the observed children when they are in fourth and fifth grade and regard those as results of the accumulated attendance of an all-day school, we explore the other waves of our panel data to evaluate the impact of attending an all-day school measured at other ages and for further outcomes.

On the one hand, we focus on time use as outcome since attending an all-day school means that more time is spent in school and less time at home (or elsewhere outside school). This could have an impact on how much time students dedicate to doing homework, doing sports, reading, or watching TV, as well as spending time with their parents. On the other hand, we also investigate how all-day school attendance affects students' well-being. Spending more time in school could improve the social interaction in the classroom, but it could also worsen students' well-being when longer school days lead to higher stress-levels.

Figure A.2 in the Appendix shows how time use of the observed children is affected in grade two, three and four.[15] For second-graders, we see that all-day school attendance leads to a significant reduction of about half an hour in the time spent on homework in second grade. This could be explained by children already doing their homework during the afternoon program in their school. There is a negative, but not significant effect on being a member of a sports club. The all-day school coefficient is also negative, yet not significant, when it comes to third-graders' likelihood to do physical exercise at least twice a week. The same applies to fourth-graders' time spent watching TV and playing computer games. Regarding fourth-graders' probability of reading at least 30 minutes per day, the all-day school coefficient is positive and significant. In sum, the pattern of these findings suggests that some substitution in time use takes place when children attend an all-day school. The positive effect on grades in German in fourth grade (see Figure A.1) is in line with more time spent on reading when attending an all-day school.

When evaluating parents' answers on time use, we find an increase of 29 ppt in the probability that parents feel that they are not spending enough time with their child when the child attends an all-day school. The effect on actual time spent together on weekdays is negative, yet not significant.

As regards students' well-being, the observed children are asked about their satisfaction with school when they are in third grade. We find a positive and marginally significant effect of all-day school attendance on school satisfaction. Importantly, children who attend an all-day school also become significantly less often victimized by bullying as reported by parents and teachers. This suggests that the social interactions in school may indeed improve due to all-day programs. Since students' well-being is an important prerequisite for successful learning (OECD 2017), these findings are well in line with the positive effect of attending an all-day school on grades shown in Figure 4.

6 Heterogeneity Analysis

Figure 4 shows the impact of attending an all-day school for the average student. It is, however, highly likely that the effect could differ by students' demographics and family backgrounds. In fact, heterogeneous effects are intended as all-day schools mainly target fostering low performing students' achievement, see Section 2. Furthermore, the counterfactual afternoon program potentially varies by family background.[16] Finally, girls and boys may differ in how well they cope with doing homework at their home,

[15]As described in Section 3.1, the NEPS data does not provide comparable items over time for our outcome variables. This is why we cannot estimate the effect of all-day school attendance on the same group of outcomes in every grade.

[16]For a discussion of the counterfactual situation of formal care in afternoon hours in Germany see Felfe and Zierow 2014 and Felfe and Zierow 2018.

instead of in a structured afternoon program, e.g., Driessen and Langen 2013.

We investigate heterogeneous effects by including an interaction term in our model which then adopts the following form:

$$y_{ism} = \beta' X_{ism} + \gamma \hat{Allday}_{ism} + \tau \hat{interact}_{ism} + \delta_b + \epsilon_{ism} \tag{3}$$

Where $interact_{ism} = Allday_{ism} \times sg_{ism}$ is the product of the all-day school attendance and a subgroup identifier sg_{ism} which is a dummy variable for being female, a non-native speaker, from a single-parent household and a household, with parents holding an academic degree. Importantly, both the all-day school attendance and the interaction term are instrumented. We include the interaction of the investment per student, and the subgroup identifier as additional instruments for the respective first stages. The results for the four subgroups are shown in Table A.5 in the Appendix.

Concerning the child's gender, the results suggest no large difference in the effect of all-day programs. The interaction is insignificant for all five outcomes.

Non-native speakers are negatively affected in terms of their test scores in math and German. The all-day program is hence estimated to widens the native vs. non-native gap in math and German (see Table 2). Also, with respect to grades and high-track attendance, non-native speakers do not seem to benefit as much from attending an all-day school as the group of native speakers, although the interaction is not significant for those outcomes.

This pattern suggests that the following scenario is not at play: One could think that non-native speakers are benefiting from all-day programs as soon as they are enrolled, even though they might be negatively affected when they attend a school with an all-day program but are not taking part in it. Then, negative effects could be explained by being excluded from additional training. However, our results show that not only attending a school with a higher all-day share[17], but actually being enrolled in an all-day program (Table A.5) yields negative effects on test scores for the group of non-native speakers.

As regards differences of the all-day school effect by parental educational background, our results suffer most seriously from a lack of statistical power. However, Table A.5 shows that the coefficients of the interaction term (high-educated parents and all-day school attendance) are negative, suggesting that achievement gaps between children from different educational backgrounds are rather diminished by all-day school programs.

In terms of children living with a single parent, the results point towards smaller

[17]Heterogenous effects for the all-day share on school level are available upon request. The results generally point in a very similar direction as the results for individual all-day attendance presented in Table A.5

effects of attending an all-day school compared to children cohabiting with both parents. Especially in terms of later high-track attendance, children of single parents are not as much benefiting from their all-day school attendance as their peers do (see Table A.5).

There are some limitations to the interpretation of our heterogeneity analysis. As already mentioned, a lack of statistical power is an issue for some of the presented estimations. Including the interaction term in equation (3) as a second endogenous variable increases the requirements for our instrumental variable strategy substantially. It changes the situation to a case with two instruments and two instrumented variables. Therefore, we report values of the Kleibergern-Paap-rank statistic for the strength of the first stage in table A.5 instead of F-values (Kleibergen and Paap 2006). Stock and Yogo (2005) present critical values which can be used for the Kleibergern-Paap-rank statistic (and the related Cragg-Donald statistic) to establish a lower bound for the significance. In our case, the critical value to have at least a significance of 90 per cent for a t-test of a coefficient on 95 per cent significance is 7.03. Estimates which are tested to be significant on only 90 per cent level or have a Kleibergen-Paap-rank statistic below the critical value (this is the case for the estimations which include the interaction with parental education) should be interpreted with caution. Their true significance level could be below 90 per cent.

In sum, our heterogeneity analysis does not provide evidence for all-day schools resulting in a reduction of inequality within the German educational system. The results rather suggest that groups of students with an already weaker performance on average, as non-native speakers and children raised by single-parents, are falling even further behind when children attend all-day schools. An initial potential explanation for these results could be the voluntary nature of activities offered to children enrolled in an afternoon program. If students can sort into voluntary activities in the afternoon hours, the sorting might not be based on selection-on-gains. If children have the choice, they potentially opt for learning activities which they are already good at than activities they would have difficulties with.[18] In such a scenario, it would not be so easily possible to harness the compensating potential of all-day schools.

7 Conclusion

In this study, we investigate whether the extension of school days induced by voluntary all-day programs in German primary schools is successful in increasing school performance. Making use of the data of the second cohort of the National Educational Panel Study (NEPS), we employ an instrumental variable approach. Our approach relies on a

[18]Unfortunately, we do not have access to any data on which kind of afternoon program activities the observed children choose. Therefore, we cannot empirically test this potential channel.

federal all-day school investment project set up to promote all-day schools by providing construction subsidies at the amount of over four billion euros in total. We exploit the information on the federal investment received by municipalities as an instrumental variable.

On average, we do not find significant effects of all-day school programs on students' standardized test scores. However, we do find a positive average effect on children's grades as well as on the probability to attend a high-track secondary school. Furthermore, our results indicate that all-day school attendance increases school satisfaction of children and decreases the probability to be bullied in school.

Our subgroup analysis reveals that non-native speakers and children with single parents do not benefit as much from all-day programs as their peers do. In sum, we find evidence that, on average, all-day school programs are beneficial for children in terms of their school grades and the important transition into the next school track. We cannot, however, provide proof that all-day programs contribute to decreasing inequality within the school system. Finding the reasons for this result would be an important topic for future research on all-day school programs, e.g. by collecting data on which afternoon activities the different subgroups of students in all-day schools sort into, or investigating the quality of all-day school programs by student subgroups.

The positive effects of all-day schools on grades and academic trajectories of students are an important finding. Both are linked to beneficial outcomes in later life. Impacting these positively would be a tangible success for all-day programs.

References

Aassve, Arnstein, Bruno Arpino, and Alice Goisis (2012). “Grandparenting and mothers' labour force participation: A comparative analysis using the generations and gender survey”. In: *Demographic Research* 27, pp. 53–84.

BBSR (2018). *Inkar: Indikatoren zur Raum- und Stadtentwicklung.* Tech. rep. Bundesinstitut für Bau-, Stadt- und Raumforschung.

Bellei, Cristián (2009). “Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile”. In: *Economics of Education Review* 28.5, pp. 629–640.

Blau, David and Janet Currie (2006). “Pre-School, Day Care, and After-School Care: Who's Minding the Kids?” In: vol. 2. Handbook of the Economics of Education. Elsevier. Chap. 20, pp. 1163–1278.

Blossfeld, Hans-Peter, Hans-Guenther Roßbach, and Jutta von Maurice (2011). “Education as a Lifelong Process: The German National Educational Panel Study (NEPS)”. In: *Zeitschrift für Erziehungswissenschaft* 14.

BMBF (2009). *Gut angelegt - Das Investitionsprogramm Zukunft Bildung und Betreuung*. Bundesministerium für Bildung und Forschung.

Bowers, Alex J. (2011). "What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school". In: *Educational Research and Evaluation* 17(3), pp. 141–159.

Brookhart, Susan M., Thomas R. Guskey, Alex J. Bowers, James H. McMillan, Jeffrey K. Smith, Lisa F. Smith, Michael T. Stevens, and Megan E. Welsh (2016). "A Century of Grading Research: Meaning and Value in the Most Common Educational Measure". In: *Review of Educational Research* 86(4), pp. 803–848.

Cannon, Jill S., Alison Jacknowitz, and Gary Painter (2006). "Is full better than half? Examining the longitudinal effects of full-day kindergarten attendance". In: *Journal of Policy Analysis and Management* 25.2, pp. 299–321.

— (2011). "The effect of attending full-day kindergarten on English learner students". In: *Journal of Policy Analysis and Management* 30.2.

Cortes, Kalena E., Joshua S. Goodman, and Takako Nomi (2015). "Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra". In: *Journal of Human Resources* 50.1, pp. 108–158.

Dehos, Fabian T. and Marie Paul (2021). "The Effects of After-School Programs on Maternal Employment". In: *Journal of Human Resources*.

Destatis (2017). *Gemeindeverzeichnis: Gebietsstand 31.12.2015*. Tech. rep. Statistisches Bundesamt.

Dobbie, Will and Jr. Fryer Roland G. (2013). "Getting beneath the Veil of Effective Schools: Evidence from New York City". In: *American Economic Journal: Applied Economics* 5.4, pp. 28–60.

Driessen, Geert and Annemarie van Langen (2013). "Gender Differences in Primary and Secondary Education: Are Girls Really Outperforming Boys?" In: *International Review of Education* 1, pp. 67–86.

Felfe, Christina and Larissa Zierow (2014). "After-School Center-Based Care and Children's Development". In: *The B.E. Journal of Economic Analysis & Policy* 14.4.

— (2018). "From dawn till dusk: Implications of full-day care for children's development". In: *Labour Economics* 55.C, pp. 259–281.

Huebener, Mathias, Susanne Kuger, and Jan Marcus (2017). "Increased instruction hours and the widening gap in student performance". In: *Labour Economics* 47, pp. 15–34.

Jackson, C. Kirabo (2018). "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes". In: *Journal of Political Economy* 126(5), pp. 2072–2107.

Kleibergen, Frank and Richard Paap (July 2006). "Generalized reduced rank tests using the singular value decomposition". In: *Journal of Econometrics* 133.1, pp. 97–126.

Klemm, Klaus and Dirk Zorn, eds. (2017). *Gute Ganztagsschule für alle – Kosten für den Ausbau eines qualitätsvollen Ganztagsschulsystems in Deutschland bis 2030*. 1., überarb. Aufl. Gütersloh: Bertelsmann Stiftung.

KMK (2015). *Allgemeinbildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland - Statistik 2009 bis 2013*. Bonn: The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany.

Kuger, Susanne (2006). "Curriculum and learning time in international school achievement studies". In: ed. by Susanne Kuger, Eckhard Klieme, Nina Jude, and David Kaplan. Assessing contexts of learning: An international perspective. Springer.

Kuhlmann, Christian and Klaus-Jürgen Tillmann (2009). "Mehr Ganztagsschulen als Konsequenz aus PISA? Bildungspolitische Diskurse und Entwicklungen in den Jahren 2000 bis 2003". In: *Ganztagsschule als symbolische Konstruktion: Fallanalysen zu Legitimationsdiskursen in schultheoretischer Perspektive*. Ed. by Fritz-Ulrich Kolbe, Sabine Reh, Till-Sebastian Idel, Bettina Fritzsche, and Kerstin Rabenstein. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 23–45.

Kultusministerkonferenz (2002). *PISA 2000 - Zentrale Handlungsfelder: Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen in den Ländern*. URL: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2002/2002_10_07-Pisa-2000-Zentrale-Handlungsfelder.pdf (visited on 09/24/2020).

Lavy, Victor (2015). "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries". In: *The Economic Journal* 125.588.

Lavy, Victor and Analía Schlosser (2005). "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits". In: *Journal of Labor Economics* 23.4, pp. 839–874.

Linberg, Tobias, Olaf Struck, and Thomas Bäumer (2018). "Vorzug Ganztagsschule?: Zusammenhänge mit der Kompetenzentwicklung im Bereich Lesen und Mathematik". In: *Zeitschrift für Erziehungswissenschaft* 14, pp. 1205–1227.

Mandel, Philipp, Bernd Süssmuth, and Marco Sunder (2019). "Cumulative instructional time and student achievement". In: *Education Economics* 27.1, pp. 20–34.

Meyer, Erik and Chris Van Klaveren (2013). "The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands". In: *Economics of Education Review* 36.Supplement C, pp. 1–11.

Nemitz, Janina (2016). "The Effect of All-Day Primary School Programs on Maternal Labor Supply". In: *Beitraege zur Jahrestagung des Vereins fuer Socialpolitik 2016.*

OECD (2011). *PISA quality time for students: Learning in and out of school.* Organisation for Economic Co-operation and Development.

— (2017). "PISA 2015 Results (Volume III)". In: OECD Publishing, Paris, France. Chap. Students' Well-Being.

Oster, Emily (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence". In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pfänder, Hanna, Michael Schurig, Martin Burghoff, and Johanna Otto (2018). "Rahmenmodell für Entwicklungsprozesse an Ganztagsschulen". In: *Journal for educational research online* 10.1, pp. 5–23.

Plantenga, Janneke and Chantal Remery (2013). *Childcare services for school age children: A comparative review of 33 countries.* Publications Office of the European Union.

Rivkin, Steven G. and Jeffrey C. Schiman (2015). "Instruction time, Classroom Quality, and Academic Achievement". In: *The Economic Journal, 125(588)*, pp. 425–448.

Sauerwein, Markus N (2019). "Partizipation in der Ganztagsschule–vertiefende Analysen". In: *Zeitschrift für Erziehungswissenschaft* 22.2, pp. 435–459.

SPI NRW (2010). *Investitionsprogramm Zukunft Bildung und Betreuung 2003-2009. Geförderte Schulen in den Jahren 2003, 2004, 2005, 2006, 2007, 2008, 2009, verschiedene Bundesländer.* Tech. rep. Sozialpädagogisches Institut NRW.

Spieß, C Katharina, Felix Büchel, and Joachim R Frick (2002). "Kinderbetreuung in West-und Ostdeutschland: sozioökonomischer Hintergrund entscheidend". In: *DIW Wochenbericht* 69.31, pp. 518–524.

StEG (2013). *Ganztagsschule 2012/2013. Deskriptive Befunde einer bundesweiten Befragung.* Frankfurt am Main, Dortmund, Gießen, München: StEG-Konsortium.

Steinmann, Isa, Rolf Strietholt, and Daniel Caro (2018). "Participation in extracurricular activities and student achievement: Evidence from German all-day schools." In: *School Effectiveness and School Improvement* 44, pp. 1–22.

Stock, James H. and Motohiro Yogo (2005). "Identification and Inference for Econometric Models Essays in Honor of Thomas Rothenberg". In: Cambridge University Press. Chap. Testing for Weak Instruments in Linear IV Regression.

A Appendix

Table A.1: Control regression

	(1) federal investment	
Chosen control variables:		
inhabitants (100'000s)	29.5***	(0.000)
female employment rate	1.467	(0.107)
tax capacity (100 Euro p.c.)	40.4**	(0.004)
Candidate control variables:		
GDP p.c.	-0.367	(0.332)
apprenticeships-applicants ratio	-0.802	(0.196)
household income	-0.00170	(0.929)
share intermediate high school exam	-0.382	(0.492)
share university entrance diploma	0.190	(0.747)
share of school leavers without exam	-1.442	(0.263)
labor market participation rate	-1.469	(0.121)
children below age 6	1.194	(0.771)
outward migration	0.0362	(0.888)
inward migration	-0.0369	(0.885)
rate of birth	0.460	(0.809)
Constant	145.4*	(0.046)
N	4461	
Test on joint significance of all candidates:		
FValue	0.759	
p	0.681	

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable is the instrumental variable, the investments of the IZBB-program in 1'000 Euro per student (as vigintile median), see Sections 4 and 3.1.

All independent variables show the values of 2003, before the investments started. State fixed-effects are included.

The variables on county level are the GDP p.c, apprenticeship-applicants ratio, the household income in Euro p.c. and the shares for different school leaving certificates/ no certificates among the school leavers. The reference category is the intermediate high school certificate.

The variabes on municipality level are the labor market participation rate and the share of children below age six, outward and inward migration and the rate of birth, all four variables are given in shares of the population.

The unit of observation is the so-called "Gemeindeverband" (roughly municipality association).

Source: BBSR 2018 and SPI NRW 2010.

Table A.2: Main results: test scores

	Maths		German	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
all-day attendance	-.0744*	.371	-.1**	.176
	(0.059)	(0.332)	(0.016)	(0.681)
individual controls:				
female	-.133***	-.135***	.218***	.217***
	(0.000)	(0.000)	(0.000)	(0.000)
non-native speaker	-.072*	-.0846**	-.0413	-.0498
	(0.059)	(0.034)	(0.241)	(0.198)
age in months	-.0156***	-.0151***	-.0102***	-.00983***
	(0.000)	(0.000)	(0.004)	(0.005)
parent, college degree	.436***	.437***	.424***	.425***
	(0.000)	(0.000)	(0.000)	(0.000)
single parent household	-.0912**	-.133**	-.0664	-.0936
	(0.046)	(0.026)	(0.160)	(0.143)
sibling in household	-.0104	.0427	-.0571*	-.0241
	(0.769)	(0.470)	(0.099)	(0.697)
school and teacher controls:				
private school	.0486	-.00299	.139*	.106
	(0.636)	(0.984)	(0.089)	(0.307)
low SES school	-.0419	-.0306	-.0656	-.0598
	(0.329)	(0.509)	(0.135)	(0.191)
share of migrants in school	-.315***	-.362***	-.17	-.198
	(0.003)	(0.007)	(0.155)	(0.160)
male teacher	.0266	.0266	-.00976	-.00809
	(0.681)	(0.674)	(0.883)	(0.902)
teacher age	.00513***	.00474***	.00464**	.00436**
	(0.003)	(0.008)	(0.012)	(0.020)
full-time teacher	-.000976	-.0154	.0394	.0294
	(0.981)	(0.718)	(0.335)	(0.482)
municipality controls:				
female employment rate	.00397	.00378	.000656	.000561
	(0.468)	(0.510)	(0.909)	(0.923)
inhabitans (100'000s)	.0256*	.0129	.0211	.0132
	(0.060)	(0.470)	(0.162)	(0.485)
tax capicity (100 Euro p.c.)	.00365	-.000415	.00826	.00559
	(0.732)	(0.971)	(0.419)	(0.606)
FSt		18.3		17.8
N	4270	4270	4361	4361

p-values in parentheses

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

The dependent variables are standardized test scores in math and German.
All-day attendance is a 0-1 identifier.
Control dummies for missing information on share of students with migrant background, SES of students and teacher's age, gender and working hours and state fixed-effects are included. Standard errors are clustered on school level.
The F-statistic for excluding the instrument in the first stage is given for the IV-estimations in the bottom panel.

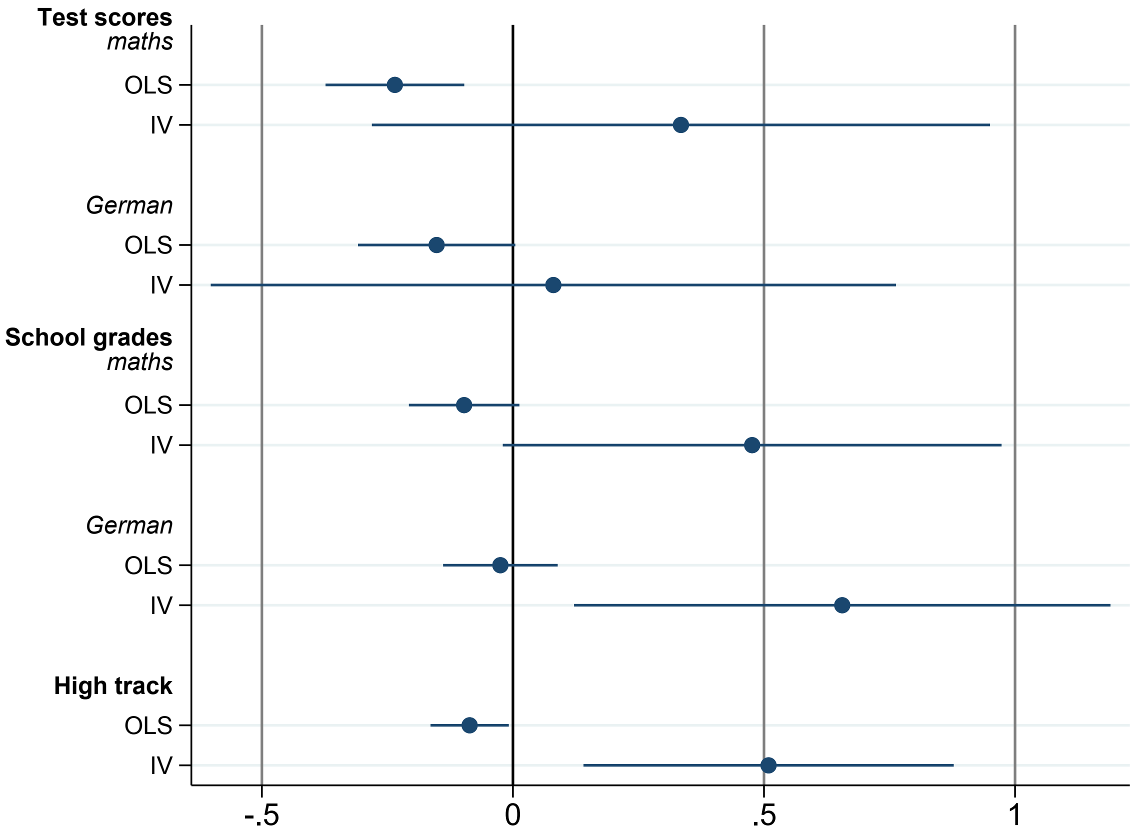
Table A.3: Main results: school grades and tracking decision

	Maths		German		High track	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
all-day attendance	-.0189	.591*	-.0035	.671**	-.0319	.48**
	(0.611)	(0.066)	(0.924)	(0.047)	(0.154)	(0.023)
individual controls:						
female	-.0724***	-.0719**	.273***	.274***	.0137	.0154
	(0.009)	(0.015)	(0.000)	(0.000)	(0.409)	(0.401)
non-native speaker	-.028	-.0511	-.018	-.0443	.0295	.0085
	(0.487)	(0.245)	(0.631)	(0.287)	(0.194)	(0.745)
age in months	-.00639*	-.00573	-.00538	-.0047	-.0058***	-.00535***
	(0.090)	(0.141)	(0.104)	(0.176)	(0.002)	(0.007)
parent, college degree	.322***	.324***	.308***	.31***	.208***	.206***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
single parent household	-.206***	-.275***	-.14***	-.218***	-.104***	-.151***
	(0.000)	(0.000)	(0.005)	(0.001)	(0.001)	(0.000)
sibling in household	.0066	.0792	-.0479	.0331	-.0234	.0351
	(0.844)	(0.122)	(0.131)	(0.521)	(0.251)	(0.305)
school and teacher controls:						
private school	-.128	-.206	-.137*	-.221*	-.00315	-.0616
	(0.108)	(0.130)	(0.066)	(0.089)	(0.950)	(0.372)
low SES school	.0105	.0271	-.0612	-.0424	-.0487*	-.0346
	(0.766)	(0.506)	(0.112)	(0.311)	(0.058)	(0.236)
share of migrants in school	.0112	-.0488	-.0236	-.0893	.151**	.0851
	(0.902)	(0.680)	(0.811)	(0.445)	(0.016)	(0.311)
male teacher	.0673	.0561	.0932*	.0802	.0544	.0504
	(0.142)	(0.242)	(0.072)	(0.151)	(0.108)	(0.189)
teacher age	-.00138	-.00242	-.0015	-.00267	-.00121	-.00214*
	(0.424)	(0.221)	(0.361)	(0.170)	(0.309)	(0.097)
full-time teacher	-.00972	-.0317	.0244	.000342	-.00536	-.022
	(0.791)	(0.440)	(0.531)	(0.993)	(0.827)	(0.441)
municipality controls:						
female employment rate	-.000817	.000641	.00189	.00344	-.0036	-.00291
	(0.866)	(0.906)	(0.721)	(0.549)	(0.287)	(0.444)
inhabitans (100'000s)	.00319	-.0124	.0137	-.00355	.00157	-.0114
	(0.751)	(0.387)	(0.181)	(0.820)	(0.816)	(0.272)
tax capicity (100 Euro p.c.)	.00686	-.00107	.00439	-.0043	.0172***	.0107
	(0.456)	(0.929)	(0.623)	(0.712)	(0.009)	(0.188)
FSt	.	23.7	.	22.7		22.1
N	3597	3597	3594	3594	3459	3459

p-values in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The dependent variables are school grades from 1 (fail) to 6 (very good) in math and German and attending the high track in high school.
All-day attendance is a 0-1 identifier.
Control dummies for missing information on share of students with migrant background, SES of students and teacher's age, gender and working hours and state fixed-effects are included. Standard errors are clustered on school level.
The F-statistic for excluding the instrument in the first stage is given for the IV-estimations in the bottom panel.

Figure A.1: The impact of all-day share in school



Notes: OLS- and IV (2SLS)-estimates for the all-day share (which ranks from zero to one) and the 90 percent confidence intervals.

The dependent variables are normalized test scores, school grades (from 1(fail) to 6(very good)) and attendance of high track in high school.

State fixed-effects and a set of individual, teacher, school, and municipality level controls and controls for missing values in those variables are included. Standard errors are clustered at school level.

Source: NEPS SC 2.

Table A.4: Sensitivity analysis

	Test scores		School grades		
	maths	German	maths	German	high track
estimates for the all-day share:					
baseline regression (see table A.2 and A.3)	.34	.152	.614*	.699**	.495**
	(0.377)	(0.727)	(0.064)	(0.045)	(0.021)
FSt	18.1	17.6	23.3	22.2	21.9
in a regression with additional controls for:					
additional classes in main subjects	.378	.141	.669*	.77*	.53**
	(0.377)	(0.759)	(0.071)	(0.051)	(0.023)
FSt	15.7	15.1	20.4	19.5	19.5
high share of qualified employees	.338	.153	.598*	.706**	.487**
	(0.371)	(0.720)	(0.070)	(0.041)	(0.021)
FSt	19.1	18.6	23.9	22.9	22.6
support for students at risk	.346	.163	.623*	.71**	.5**
	(0.380)	(0.712)	(0.064)	(0.049)	(0.023)
FSt	17.2	16.8	23.6	22.4	21.2
all three variables above	.378	.152	.654*	.786**	.524**
	(0.372)	(0.738)	(0.077)	(0.049)	(0.025)
FSt	16	15.4	21	19.9	19.6
N	4270	4361	3597	3594	3459

Standard errors in parentheses

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

The dependent variables are normalized test scores, school grades (from 1(fail) to 6(very good)) and attendance of high track in high school.

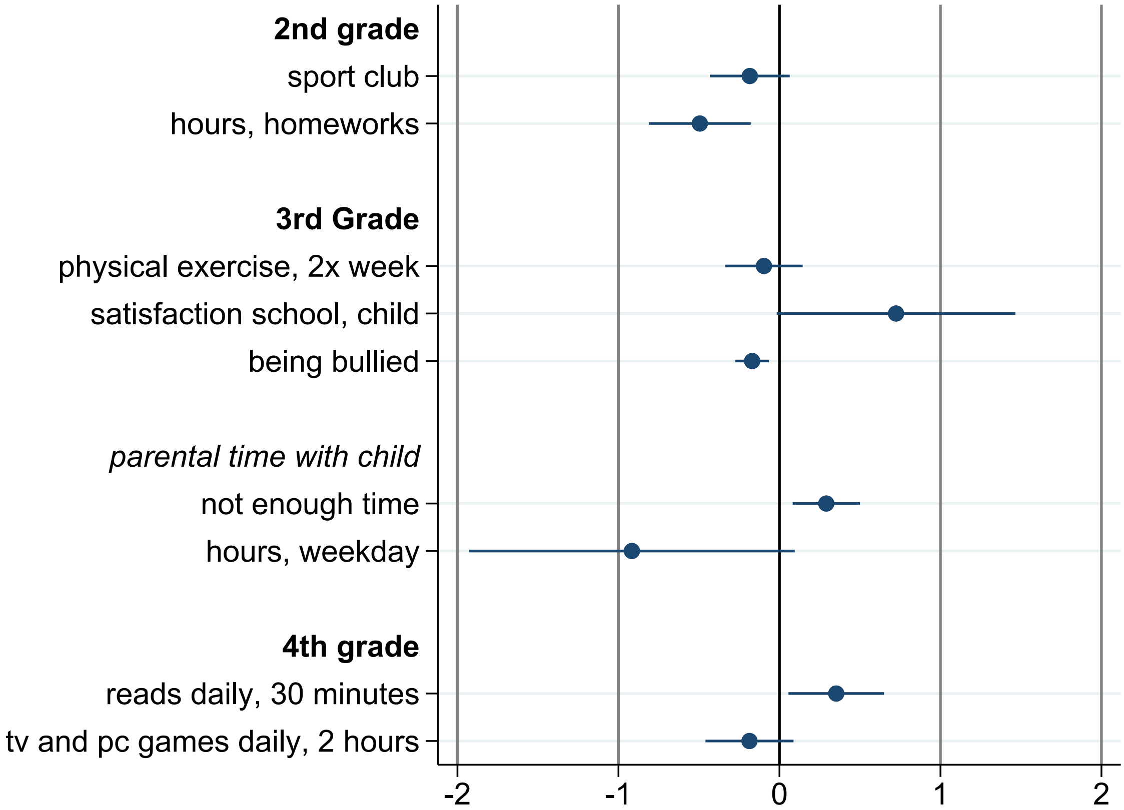
The table shows the estimates for the share of all-day students in the baseline, in a regression with the respective additional control and in a regression controlling for all three auf those variables.

The additional classes-dummy is one for schools which offer additional clases in math, German and science. The high share of qualified employees-dummy is one for schools employing more people qualified in child care or teaching than the sample mean. The dummy for offering support for students at risk is one for schools offering support for low-performing students and non-native speakers.

For the distribution of these three variable see table 1.

State fixed-effects and the full set of control variables shown in tables A.2 and A.3 on individual, teacher, school and municipality level controls and controls for missing in those variables are included. Standard errors are clustered on school level.

Figure A.2: The impact of individual all-day attendance on further outcomes



Notes: Estimates for individual all-day attendance and the 90 percent confidence intervals.
For 2nd grade, the dependent variables are a dummy for being member in sport club and the time spent on home works in hours. Both answers are given by the parents.
For 3rd grade, the dependent variables are dummies for doing physical exercise at least twice per week, having a high general satisfaction with the school, answered given by the child and for being the victim of bullying according to the parent or the teacher. Further, there is a dummy for not spending enough time with the child and the time spent with the child in hours on weekdays, answered by the parents
For 4th grade, the dependent variables are dummies for reading at least 30 minutes per day and for spending at least two hours on TV and PC-games. Both answers are given by the child.
State fixed-effects and the full set of control variables shown in tables A.2 and A.3 on individual, teacher, school and municipality level controls and controls for missing in those variables are included. Standard errors are clustered on school level.
Source: NEPS SC 2.

Table A.5: Heterogeneity analysis, individual all-day attendance

	Test scores		School grades		
	maths	German	maths	German	high track
Interaction with female					
individual attendance	.481	.263	.563*	.568*	.543***
	(0.250)	(0.531)	(0.083)	(0.094)	(0.007)
interaction term	-.265	-.215	.0964	.254	-.0901
	(0.394)	(0.438)	(0.680)	(0.290)	(0.527)
KPrkSt	8.83	8.52	11.5	11	10.9
Interaction with indicator for non-native speakers					
individual attendance	.485	.31	.664**	.726**	.541**
	(0.219)	(0.499)	(0.045)	(0.038)	(0.016)
interaction term	-.715**	-.765**	-.292	-.161	-.272
	(0.048)	(0.021)	(0.356)	(0.540)	(0.119)
KPrkSt	9.15	8.91	11.7	11.2	11
Interaction with indicator for parents with university diploma					
individual attendance	.374	.0722	1.23**	1.1*	.715*
	(0.546)	(0.911)	(0.037)	(0.096)	(0.059)
interaction term	-.0381	.09	-.697*	-.445	-.262
	(0.924)	(0.817)	(0.084)	(0.310)	(0.342)
KPrkSt	4.28	4.35	6.57	6.07	5.82
Interaction with indicator for single-parent household					
individual attendance	.337	.166	.61*	.728**	.543**
	(0.398)	(0.713)	(0.071)	(0.044)	(0.022)
interaction term	.0399	-.181	.0512	-.47	-.467**
	(0.913)	(0.638)	(0.888)	(0.193)	(0.039)
KPrkSt	8.77	8.59	11.6	11	10.1
N	4270	4361	3597	3594	3459

p-values in parentheses

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

The dependent variables are normalized test scores, school grades (from 1(fail) to 6(very good)) and attendance of high track in high school.

KPrkSt: Kleibergen-Paap-rank statistic is the measurement for the strength of the first stage (comparable to the F-statistic in the case with one instrument and one instrumented variable)(Kleibergen and Paap 2006). Low values indicate weak instruments and that the true significance of a t-test for the second stage's coefficients could be lower than intended by the test. A test on 95 significance has at least 90 significance if the Kleibergen-Paap-rank statistic exceeds 7.03 (Stock and Yogo 2005).

The interaction term is constructed as product of individual all-day attendance and the respective subgroup-identifier.

State fixed-effects and the full set of control variables shown in tables A.2 and A.3 on individual, teacher, school and municipality level controls and controls for missing in those variables are included. Standard errors are clustered on school level.