

Seidlitz, Arnim; Zierow, Larissa

Working Paper

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

CESifo Working Paper, No. 8618

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Seidlitz, Arnim; Zierow, Larissa (2020) : The Impact of All-Day Schools on Student Achievement - Evidence from Extending School Days in German Primary Schools, CESifo Working Paper, No. 8618, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<http://hdl.handle.net/10419/226320>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

Arnim Seidlitz, Larissa Zierow

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>

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. We combine data from the National Educational Panel Study covering 5771 primary school students with municipality-level information on all-day school investments. 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. Results imply that all-day programs lead to improvements in language and math skills as measured by teacher assessments and to a higher probability of being recommended for the academic track after primary school. The heterogeneity analysis reveals that boys benefit more than girls from all-day programs in terms of the assessment by their math teacher. Furthermore, there is a significant negative effect on non-native speakers' math and German test scores.

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

October 2020

We thank Eric Hanushek, Ludger Woessmann, Bernd Fitzenberger and seminar participants at HU Berlin, ifo Institute, the 4th international NEPS Conference, the EALE-SOLE-AASLE World Conference 2020 and the EEA Congress 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 1st to 4th grade. 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 as a vehicle for increasing equality of opportunity in the education system. An increase in the time students spend at school can provide extended learning opportunities and increased individual support to foster the academic and psychosocial development of children, especially those from more disadvantaged backgrounds (Plantenga and Remery 2013; Blau and Currie 2006). In a system of half-day schools, up to 38% of the total weekly learning time takes place in 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 can become important drivers of inequality in learning opportunities. Therefore, the potential to enhance equality of opportunity by regulating the quantity and quality of afternoon learning activities for all children is one major advantage of all-day programs (Kuger 2006).

This line of argument led many governments to introduce more school-based afternoon activities over the last 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 badly in this first international student assessment of the OECD – both in terms of average achievement and equality of opportunity – politicians and experts were looking for suitable reforms that could improve the school system. The introduction of all-day schools was one of the measures taken (Kultusministerkonferenz 2002).

In this paper, we investigate the extension of all-day school programs in Germany which was triggered by a large investment program by the federal government. 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 further school performance measures. We additionally investigate the impact on children's time use and school satisfaction. Importantly, in addition to evaluating the mean 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 making use of the exogenous nature of a funding program by the German federal government. Notably, the funds (4 billion euros in total) could be used for all-day school related constructional purposes only.¹ The program was quite successful as the share of all-day school students has been rising dramatically ever since. In 2002, 1% of primary school children attended an all-day school, compared to 22% in 2012 (Dehos and Paul 2017).

Our analysis uses very detailed administrative data of the federal program containing the

¹Since the different German states are in charge of school affairs (e.g. hiring teachers), the federal government is restricted in the allocation of funds to the educational system.

amount of 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 5771 primary school students, including our outcomes of interest. The funding received by the federal funding program serves as instrument for the share of all-day school students in every school contained in NEPS. The first stage results show that the instrument is strongly associated with the all-day school share. We furthermore 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 have a positive effect on children’s achievements as assessed by their classroom teacher: a 20 percentage point increase in the all-day share leads to a 12 percentage point higher probability to be evaluated as (at least) ”good” in math and German. Such an increase would also lead to an eight percentage point higher probability to be recommended for the highest track. We do not, however, find significant effects of all-day school programs on children’s scores in standardized math and German tests. Evaluating the effect of all-day schools 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. The heterogeneity analysis reveals that boys benefit more than girls from all-day programs in terms of the assessment by their math teacher. Furthermore, non-native speakers are significantly negatively affected in terms of their test scores in math and German. We therefore cannot provide proof that the all-day programs help much in decreasing inequality in the school system.

Our study relates to a growing literature investigating the impact of instructional time on student outcomes. Findings from these studies are mixed. On the one hand, several studies find positive effects of instructional 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). On the other hand, Cannon et al. 2011 and Meyer and Klaveren 2013 find no or only temporary gains. In terms of educational inequality, Huebener et al. 2017 find that the performance gaps between students widens due to increased instructional time. If, however, additional learning input is targeted at low achieving students, educational inequality can be decreased (Cortes et al. 2015; Lavy and Schlosser 2005). We add to the debate whether all-day schools are capable of enhancing equality of opportunity in the education system (Bellei 2009; Linberg et al. 2018; Steinmann et al. 2018).

The remainder of this study is structured as follows: Section 2 describes the primary school system in Germany as well as the investment program for the expansion of all-day schools. Section 3 presents the empirical framework. Section 4 describes the data and provides descriptive statistics. 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 meaning that each of the country's 16 states is responsible for the education of its youth. Although there are some differences across states, the general structure is fairly uniform. The school system divides children at age 10, after four years of primary school, into three secondary school tracks: basic track (five years), middle track (six years), and high track (nine years). The latter leads to the university entrance qualification.

Primary schools, the school type under study, are thus very important for the students' future educational paths as children are tracked based on their performance in fourth grade. Teachers of fourthgraders 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 teacher recommendations for the secondary school track as one of our outcomes, we can shed light on the potential of all-day schools to change students' educational opportunities.

School day length in German primary schools

Until recently, the typical primary school day has been relatively short in Germany. From grade 1 to grade 4, children have instruction time starting at 8 am in the morning and finishing in the very early afternoon. As a result, the main care provider for primary school-aged children in the afternoon hours has been the mother. For instance, 64% of all primary-school-aged children were looked after by their mothers in 2011 and the majority of mothers was either not participating in the labor market or working part-time.³ In some states, after-school care centers providing care for primary school children in the afternoon hours have been implemented (supervised by the Child and Youth Services, i.e. not related to school management), but only between 5% and 20% of children attended these centers in 2005 (Felfe and Zierow 2014).

After the PISA shock in 2001, the importance of a nationwide introduction of all-day schools was getting 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 also 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.⁴ In the end, it was left to the local authorities and the schools to decide whether to introduce a mandatory or voluntary all-

²In 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 of choosing the child's secondary school track from the (limited) set of available school tracks. In our analysis we include state fixed effects, i.e. any differences regarding school track recommendations across states are controlled for.

³Calculations by Felfe and Zierow 2014 based on the German Socio-Economic Panel 2011.

⁴For an overview of the debate see Kuhlmann and Tillmann 2009.

day program. Most of the schools opted for the latter, i.e. they did not extend the compulsory instruction time but established a voluntary afternoon program.⁵ Table 1 shows the differences between half-day schools and all-day schools in terms of their afternoon activities, staff employed in these and student composition in 2014. It shows that half-day schools also provide some activities in the afternoon hours, but they do so to a much lesser extent than all-day schools. 95,5% of all-day schools provide lunch to their students. This is the case in only 62,4% of half-day schools. Similarly, homework supervision, sport activities and other activities are offered much more often by all-day schools. When it comes to supporting learning activities for specific groups of students, 84% of all all-day schools provide support for low-performing students and about half of the all-day schools have specific support for high-performing students and non-native speakers, respectively. There is also more additional teaching taking place in all-day schools. Interestingly, however, in terms of employed staff in the afternoon, all-day schools more often employ staff educated in childcare and less often staff educated in teaching.

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)*⁶ between 2003 and 2009, the federal government tried to incentivize the states to invest in the construction and expansion of all-day schools. The aim was to guarantee a nationwide supply of all-day school programs. The total investment volume was about 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. The states and local authorities therefore had to cover the personnel costs, e.g. teacher salaries, and operating costs, e.g. expenses for lunch, of the all-day school expansion. The costs for the municipalities are immense: The personnel costs for one additional all-day primary school student are estimated between 992 Euro and 1981 Euro per year, depending on the intensity of the all-day program (Klemm and Zorn 2017).

The federal funds were transferred to more than 8200 schools which 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 1% to 22% within ten years after implementation of the investment program (Dehos and Paul 2017).

Yet, the investment allocation was not uniform across municipalities and states. Therefore, the expansion of all-day slots varies a lot across regions. However, we find that pre-expansion municipality characteristics (such as birth rate, inward migration, and labor market participation rates) do not predict whether a municipality received funding (see Table A.1 in the Appendix),

⁵In our analysis we do not differentiate between mandatory and voluntary all-day programs. Note that the large majority of schools in our sample (90%) had voluntary all-day programs.

⁶*Investitionsprogramm Zukunft Bildung und Betreuung*, for more information see <https://www.ganztagsschulen.org/de/868.php>

Table 1: Contents and employees in afternoon-programs

	half-day	all-day	difference	t-statistic
school type	.471	.529		
lunch	.585	.944	.359	36.033
activities offered at least once a week				
home work supervision	.599	.93	.331	32.729
sport activities	.585	.822	.238	20.634
artistic activities	.67	.823	.154	13.688
mechanical activities	.346	.601	.255	20.021
computing activities	.373	.555	.182	14.063
support for				
high performing students	.323	.494	.171	13.398
low performing students	.638	.823	.185	16.268
non-native speakers	.449	.548	.099	7.58
additional teaching in				
maths	.212	.339	.127	10.806
science	.293	.461	.167	13.261
German	.283	.512	.229	18.201
foreign languages	.229	.283	.054	4.696
employees in afternoon programs				
mean percentages for groups				
no related qualification	.573	.576	.003	.335
qualified in childcare	.248	.286	.039	5.187
qualified in teaching	.179	.139	-.041	-7.839
share of students attending school with				
support for students at risk	.424	.523	.098	7.487
additional classes in main subjects	.075	.246	.171	17.926
a high share of qualified staff	.2	.492	.292	24.257
N	2721	3050		

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 else.

The bottom panel shows the school-subgroup-identifier we calculated from the values above and used in our sensitivity analysis, see Section 5.3.

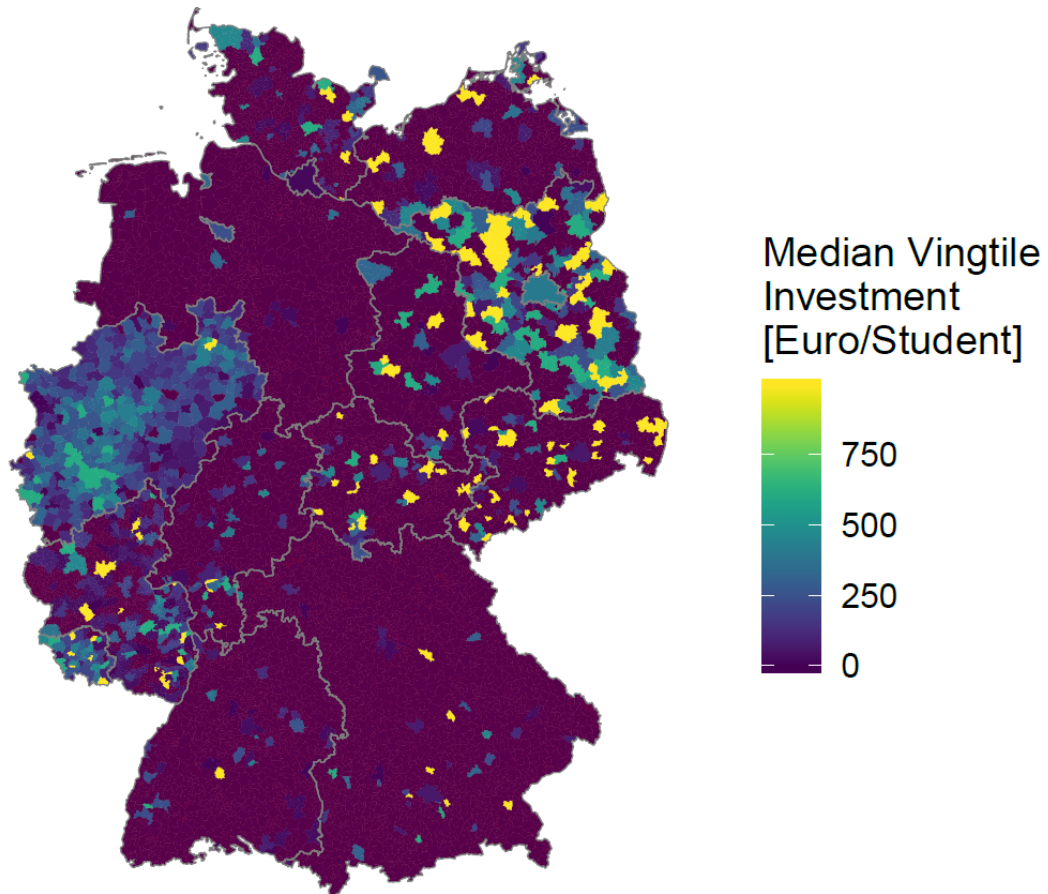
Source: NEPS SC2, own calculations based on answers from the principal's questionnaire of wave 4 conducted in 2014.

alleviating potential concerns of self-selection of municipalities into the program. A potential reason for the regional variation is the application procedure which has sometimes been slowed down by 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.⁷ Figure 1 shows the resulting regional

⁷In 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 also had to be approved by the parent council at the school conference (see Nemitz 2016 for a very comprehensive summary of the application procedure).

variation of the received funds of the federal investment program. We use this variation for our empirical strategy.

Figure 1: The all-day school investment program and its regional variation



Federal all-day school investements (SPI NRW 2010) per student (relative investments) on municipality level given as vingtile median according to the distribution of relative investments weighted by the number of inhabitants (BBSR 2018; Destatis 2017)

3 Empirical Strategy

Our analysis aims to estimate the causal impact of attending an all-day school on student achievement. Our treatment variable is the share of students in a given school who attend an all-day program, i.e. the *all-day share*. 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. A student who herself does not attend the afternoon program will be influenced by the existence of that program if a sufficiently large number of classmates does. These peer effects would be neglected if the treatment status was defined on the individual level. Since there

is also some variation in the share of all-day slots across all-day schools⁸, we decided against a binary treatment status on school level.

The main challenge of our empirical analysis is that the implementation of all-day programs did not happen at random. Local authorities with a higher share of female employment, a higher share of disadvantaged children or a higher GDP per capita may have been more in favor of expanding all-day school programs due to higher needs of families and/or larger budgets.

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).⁹ Local authorities would presumably only apply for the funding if they observed the need for implementing an all-day school in their constituency and additionally had the fiscal capacity to bear the implied costs. To account for this self-selection, we control for the share of low SES children, the share of children with an immigrant background, the size of the municipality, the local female labor market participation rate, GDP per capita and the fiscal capacity of the municipality.

For our instrumental variable, we compute the all-day school investment per student at the municipality level.¹⁰ As Figure 1 shows, we observe a large variation of this variable across municipalities ranging from 0 to 993 Euro per student.

Our first stage estimation thus looks as follows:

$$Allday_{sm} = \pi' X_{ism} + \alpha Investment_m + \delta_b + \eta_{ism} \quad (1)$$

The share of all-day students in school s in municipality m – $Allday_{sm}$ – is regressed on the instrumental variable (received investment funds in municipality m) $Investment_m$, 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 share of all-day students. For simplicity, we assume that this effect is linear.

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) the funding was only allowed to be used for constructional expenses and not for spending on teachers' salaries and other important

⁸As the expansion of all-day schools had a rather gradual nature and the attendance to all-day programs is not mandatory, there are only very few schools with 100% of students being enrolled in an all-day program.

⁹Reassuringly, the same investment program has been used as an instrumental variable for all-day school attendance by Dehos and Paul 2017 and Nemitz 2016 in their studies on the impact of all-day schools on female employment.

¹⁰For the purpose of merging the student achievement data with the investment data, we have to reduce the informational content of the investment variable due to data protection policies. Therefore, we define the instrumental variable as vingtile median according to the distribution of per-student investments weighted by the number of inhabitants.

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 got the funding, and this type of mobility is very low in Germany.

Finally, our second stage is the following:

$$y_{ism} = \beta' X_{ism} + \gamma \widehat{Allday}_{sm} + \delta_b + \epsilon_{ism} \quad (2)$$

Thus, the student achievement outcome y_{ism} of student i attending school s in municipality m is regressed on the instrumented all-day share of school s in municipality m \widehat{Allday}_{sm} , while controlling for individual, school and municipality characteristics X_{ism} as well as for state fixed-effects δ_b . The coefficient γ shows the effect of the all-day share in school s – following a LATE interpretation, the effect of the all-day share which is due to the federal investment program – on student i 's achievement.

4 Data and Descriptive Statistics

4.1 Data

To evaluate the impact of all-day schools on student achievement, we rely on three data sources. 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.¹¹

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 6000 participants in more than 300 schools are sampled at primary school entrance in 2012. Our study includes four waves of this cohort covering their four years in primary school.¹² Each wave consists of surveys for parents, class teachers, school principals and children themselves (from third grade on). Additionally, for each

¹¹For the merge with the NEPS data, we have to categorize the instrumental variable and the municipality characteristics rather than taking the continuous value in order to align with the NEPS data protection rules (disabling us to identify any municipality through our matched information).

¹²For a sub-sample of observations information on their time in kindergarten and high school would be potentially available as well. However, conditioning on the availability of information from kindergarten and high school would add selectivity and reduce the sample size below 1500 and 5000, respectively. Therefore, we decided to restrict us to the primary school outcomes. We only excluded cases with missings in all outcome variables or the municipality of residence as we are unable to match the instrument without the latter. This leaves us 5771 observations.

wave the NEPS researchers designed standardized tests in different school subjects the surveyed children had to take (Blossfeld et al. 2011). These standardized tests are a main advantage of the NEPS and an important element of our study.¹³ In fourth grade, the students are tested in orthography, reading and math. These test scores at the end of primary school serve as main outcome variables in our regressions.¹⁴

The teachers of the NEPS-participants are also asked to evaluate the performance of their students in different categories. Their subjective assessments of students' achievement may capture non-cognitive skills which are known to be important predictors for later life outcomes (Jackson 2018; Brookhart et al. 2016; Bowers 2011), but which are not captured by standardized tests. We therefore include teachers' assessment of the math and German skills of their fourth-grade students as outcomes in our analysis.

Another outcome variable is the teacher's tracking recommendation for high school surveyed in the parental questionnaire in grade 4. There is some heterogeneity concerning school types after primary school across German states. However, all German students are tracked at the end of primary school and high-track schools ("Gymnasium") exist in all states. In our regressions, we use this outcome as a dummy variable indicating whether the student was recommended for the high track by her teacher.¹⁵

From the different questionnaires, we gain further valuable control variables at the individual, teacher, school and municipality 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 living together with siblings. We further include the student's age in months. At the teacher level, we include dummies indicating the teacher to be male, to be younger than 35 years old and to work full-time. A teacher's age is our proxy for her teaching experience. At the school level, we control for being a private school, for having more than 33% of students with an immigrant background and for having more students from low socio-economic status (SES) than from high SES.¹⁶

¹³Note that the NEPS is the first (and only) data source providing nation-wide panel data on student achievement in Germany.

¹⁴In third grade, the students take part in tests for vocabulary and science. In an additional analysis we also evaluate how those test score are affected by attending an all-day school. Having student achievement data in two points in time we could theoretically perform a value-added analysis. Yet, we are restricted by having information on the instrumental variable and the endogenous variable only for one point in time. Thus, we cannot use the time dimension of the student data.

¹⁵Making use of further information provided in earlier waves of the panel we also look at outcomes beyond student achievement in a further analysis. Using data of wave 2 when students are in second grad, we evaluate whether attending an all-day school influences enrollment in a sport club and time spent on homework after school. Using data of wave 3 when students are in third grad, we evaluate whether attending an all-day school influences whether they regularly do physical exercise and whether they are satisfied with their school. Furthermore, we use parental assessment on how much time they spent with their child as outcomes. We also analyze the effect on test scores in math and German and the socio-emotional development as documented by the SDQ (Goodman 1997) when the observed children are in third grade. Finally, we assess whether attending an all-day school has an impact on time spent on reading and time spent on watching TV when children are in fourth grade.

¹⁶The school principals are asked in their questionnaire to provide an estimate on these two categories of demographic composition.

Investment Program Data and Municipality Characteristics

The 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 the year of spending for each school which 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. Approximately one quarter of the 4500 German municipalities (Gemeindeverbände) received positive funding, see Figure 1.

Furthermore, we use municipality level information on fiscal capacity, female labor market participation rate, the share of people aged 6 to 18 years as well as the GDP per capita on county level, all provided by the INKAR-data (BBSR 2018). The share of people aged 6 to 18 (BBSR 2018) and the number of inhabitants (Destatis 2017) gives us the population in this age group which is our proxy for the number of students in this municipality. As a further control, we use the number of inhabitants (Destatis 2017). For all of these municipality controls we use data from 2003, i.e. before the investments took place.

In order to merge the investment data with the student achievement data we have to minimize the investment data's informational content. Therefore, we 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 percent (vingtiles). For each of these vingtile-cells, we calculate the median investments. This median investment serves as our instrumental variable.¹⁷

4.2 Descriptive Statistics

Table 2 shows the descriptive statistics of individual and family characteristics (Panel A), school characteristics (Panel B), teacher characteristics (Panel C), municipality characteristics (Panel D), as well as the outcome variables (Panel E). The first column shows the 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, respectively (note that for this table we define all-day schools as schools having a non-zero all-day share). The right hand part of Table 2, columns 5 to 8, give the differences between subgroups of students. They show 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.

Regarding the outcome variables, there are no significant differences between the students of all-day and half-day schools. Concerning the subgroups, we see some strong but not very

¹⁷A similar procedure applies to the municipality-level control variables (female labor market participation, fiscal capacities, inhabitants and GDP per capita (county level)). In order to link them to the student achievement data, we needed to form categorical variables. For more details, see the notes of Table 2.

Table 2: Descriptives

	entire sample	all-day	half-day	t-statistic	differences according to subgroups			
					female	non-native speaker	university diploma, parents	single parent
Individual controls								
female	.511	.503	.521	-1.048		.02	-.015	.01
non-native speaker	.17	.175	.165	1.427	.011		-.064***	.025
university diploma, parents	.574	.563	.587	-.954	-.014	-.11***		-.246***
single parent	.112	.12	.104	2.079	.004	.018	-.101***	
age at school entry (months)	77.331	77.069	77.626	.578	-.747***	-.165	-.913***	.573***
living with siblings	.701	.698	.705	-1.805	.008	-.043***	.129***	-.275***
School level controls								
private school	.043	.045	.041	1.508	.005	-.015**	.019***	-.006
low SES school	.336	.349	.322	.437	-.012	.006	-.07***	.034*
migrant share $\geq 1/3$.224	.281	.16	6.875	-.004	.034**	-.072***	.029*
Teacher controls								
male teacher	.061	.067	.054	2.802	-.004	-.015*	-.01	-.007
teacher \leq age 35	.122	.115	.13	.607	-.002	.021*	-.016*	-.001
full-time teacher	.281	.297	.262	2.122	-.003	.02	-.014	-.005
Municipality controls								
GDP pc. \geq med., county	.447	.448	.445	4.403	-.002	.055***	.025**	.002
female labor market participation	44.233	43.524	45.028	-.455	-.049	.118	.108	-.068
inhabitans, in 100'000	1.2	1.559	.796	7.445	-.034	.218***	.07*	.107*
fiscal capacity	4.758	4.843	4.663	4.511	.009	.2***	.158***	-.048
Outcomes								
math test scores	.059	.042	.078	-1.462	-.118***	-.152***	.505***	-.248***
German test scores	.047	.006	.094	-8.67	.217***	-.078**	.482***	-.222***
recommendation high track	.661	.654	.67	-.393	.033**	-.007	.228***	-.118***
math teacher assessment	.501	.502	.499	-.158	-.134***	-.028	.196***	-.142***
German teacher assessment	.365	.357	.373	-.429	.095***	-.048**	.218***	-.142***
N	5771	3050	2721					

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, living with siblings in the same household and a variable controlling for age at school entry in months. The school level control variables are dummies indicating whether it is a private school, whether the school has more students from low SES than from high SES and whether there are more than a third of students having an immigrant background. Both numbers are given by the principal. The teacher controls are dummy variables indicating whether the teacher is male, at most 35 years of age and working full-time. The municipality controls are categorical variables for the GDP per capita on county level (whether GDP per capita is above state median, 2 categories), 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, 10 categories) and the fiscal capacity (rounded to 300 Euro per capita, 5 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), a dummy for being recommended by the teacher for high track and dummies for being perceived as (at least) "good" in math and German. The latter is constructed from three question for written skills, understanding and oral command.

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 (eg. 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 10, 5 and 1 percent level in a regression controlling for state fixed effects.

surprising patterns. While girls do worse in math than boys, they outperform them in language skills and are more often recommended for high track. Children of parents with a university degree show generally above average results and children raised by a single parent are below the average. Non-native speakers do worse than native speakers in standardized tests.

For the individual characteristics, we find students in all-day schools more often to live in households with only one parent and without siblings. This could be an indication that some parents use their limited opportunities for school choice – for instance avoiding the assigned public school by sending their child to a private school –, and that local authorities increased the all-day share especially in schools with a higher number of single-parents and parents of only

one child, who may have a higher demand for institutional care.

Additionally, we see that all-day schools are more often in larger municipalities (in terms of inhabitants) and in areas with a high share of students with a migrant background. This also very likely reflects the decisions of authorities where to provide all-day programs.

5 Results

In this section we present the results of our IV estimations. We begin with the first stage. Then we show the second stage results for the full sample. The section continues with a sensitivity analysis and the analysis of the effect of all-day school attendance on further outcomes.

5.1 First stage results

Table 3 shows the estimates for the first-stage regression (see Equation 1). The main variable of interest, the share of all-day students in a school, is regressed on the all-day investments per student (only targeted at constructional purpose) and the control variables. Most importantly, we find that the instrument is strongly significant and economically meaningful. An increase in the investment of 100 Euros per student is estimated to increase the share of all-day students by 2.45 percentage points. With an F-statistic larger than 14, we do not encounter a weak instrument problem.

When it comes to the control variables, we see some significant coefficients. Students without siblings and students from a single-parent household tend to attend schools with a higher all-day share. The share of all-day students is also higher in counties with an above median GDP per capita and in larger municipalities. These municipality controls contain 2003 data when the investment started, avoiding a reverse causality problem. Taken together, the estimates of the controls are well in line with the raw differences of means shown in Table 2.

5.2 Second stage results

Figure 2 shows the estimates for the all-day share coefficient in our regressions of Equation (2). It displays the impact of an increase in the all-day share on fourth-graders' test scores in math and German, on the assessment of their teacher in math and German, and on being recommended for the high track after primary school. For a more detailed regression output see Table A.2 in the Appendix.

The effects on the test scores are insignificant. Interestingly, however, there are positive and significant effects of the all-day share on the teacher assessment variables as well as on the students' high-track recommendation. The estimate for the effect on the high-track recommendation is 0.414. That means that an increase in a school's all-day share by ten percentage points

Table 3: First Stage: Regression on the share of all-day students

	(1)	
	share of students in all-day program	
investment	.000245***	(.000065)
female	-.005979	(.005048)
non-native speaker	.008165	(.008082)
age in month	.000404	(.000658)
parent, college degree	.008611	(.008291)
single parent household	.02539***	(.009741)
sibling in household	-.02054**	(.008303)
private school	.1334	(.1313)
low ses school	-.01687	(.03093)
share of migrant background $\geq 1/3$.03027	(.03857)
male teacher	.000825	(.03298)
teacher younger than 35	.02197	(.02566)
full-time teacher	.03379	(.02159)
GDP p.c. \geq median, , county	.05291*	(.02719)
female employment rate, munc.	-.000207	(.003093)
inhabitans, munc	.01947*	(.01086)
fiscal capicity, munc.	.007401	(.006336)
Constant	-.143	(.1274)
N	5771	
Fstatistic (of the instrument)	14.33	

Standard errors in parentheses

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

The share of students in an all-day program ranks from 0 to 1.

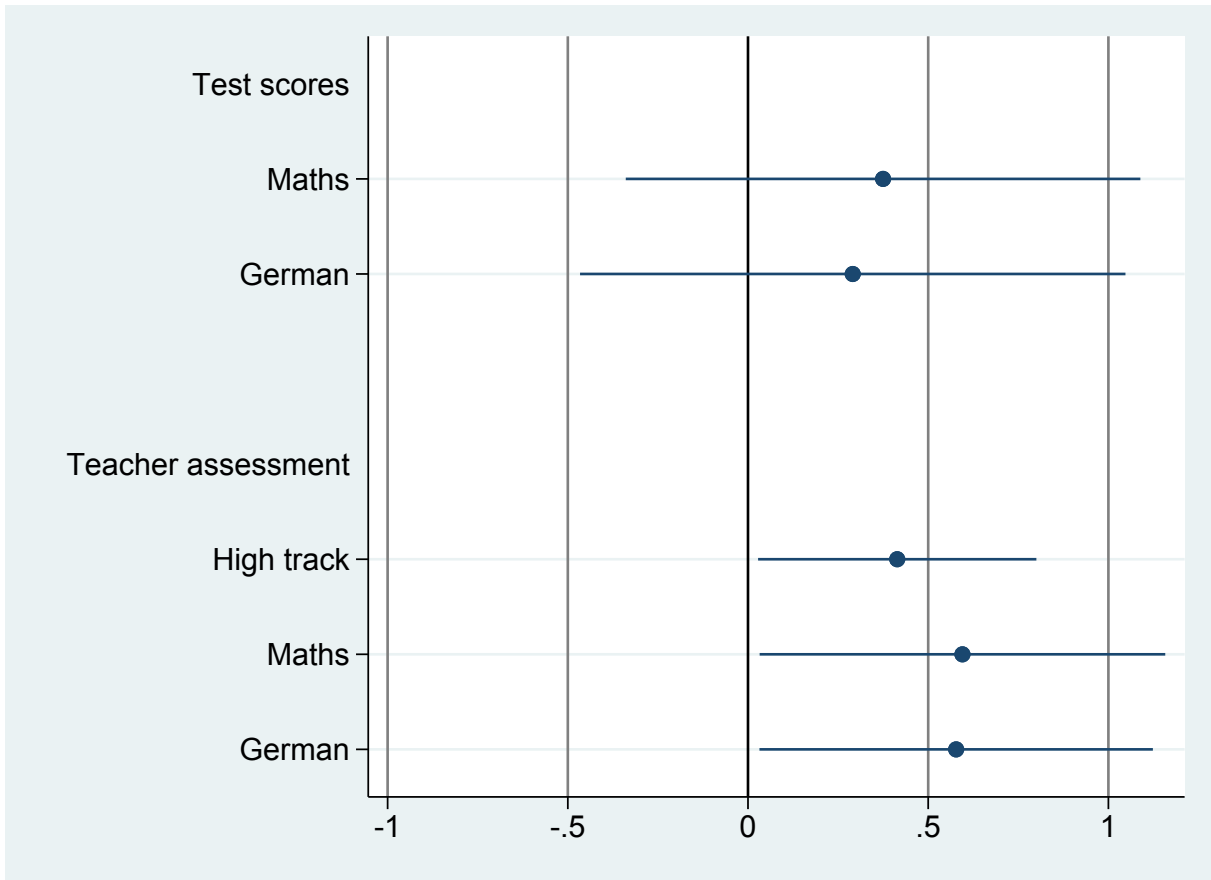
The bottom line gives the F-statistic of a test for exclusion of the instrument.

For a more detailed explanation on the control variables, see notes of table 2. Control dummies for missing information on native language, existence of siblings, organization of school, share of students with migrant background, SES of students, inhabitants of municipality and teacher's age, gender and working hours and state fixed-effects are included. Standard errors are clustered on school level.

increases the probability for a student to receive a high-track recommendation by approximately 4 percentage points. The effects on the teacher's assessment of a student's skills in math and German are significantly positive as well. An increase in a school's all-day share by 10 percentage points would result in a 6 percentage point increase of the chance to be perceived as (at least) good.

We thus find a positive effect on the subjective teacher assessments while the effect on the test scores is insignificant. This result might seem puzzling but is not self-contradictory. The perceived achievement by the teacher 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 rather have an impact on those skills which are important for the teacher assessment 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,

Figure 2: The impact of all-day school programs on student achievement



Notes: 2SLS estimates for the all-day share (which ranks from 0 to 1) and the 90 percent confidence intervals. The dependent variables are normalized test scores and dummies for teacher assessment. The dummies indicate (very) good skills compared to the average student in the respective field and a recommendation to attend the high track in secondary school.

State fixed-effects and a set of individual, teacher, school and municipality level controls and controls for missing in those variables are included. Standard errors are clustered on school level. See Table A.2 for more detailed results.

Source: NEPS SC 2.

and Bowers 2011.

5.3 Sensitivity Analysis

As shown in Table 1 in Section 2, not only all-day schools 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 could downward bias the effect of attending an all-day school. In Table A.3 in the Appendix we include the 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 controlling for any afternoon program characteristics. In the

second row, we control for 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. I.e., attending a school with a higher all-day share is more beneficial for student achievement than attending a school with a lower all-day share also when holding additional afternoon classes in both types of schools constant. The same holds true when accounting for the share of qualified employees (third row) and for support for students at risk (fourth row). When controlling jointly for all three types of afternoon programs, the respective coefficients of the treatment variable are largest, as shown in the fifth row. As regards the effects on standardized test scores the coefficients get a bit larger but are still statistically insignificant. In contrast, the effects on the other three outcomes increase when adding the quality controls and stay significant.

These results indicate that the effects of all-day schools shown in Figure 2 cannot be replicated by a system of half-day schools offering certain elements of afternoon programs. The gains seem to be exclusively related to the institution of all-day schools.

5.4 Further Outcomes

As our main outcomes we focus on test scores, teacher assessments and high-track recommendation of the observed children when they are fourth-graders. These outcomes can be regarded as the result of the accumulated attendance of an all-day school.¹⁸ 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. Figure A.1 in the Appendix shows how student achievement in third grade is affected by attending a school with a higher all-day share. Furthermore, it shows for grade two, three and four how time use of the observed children is affected.¹⁹

The test scores (in vocabulary and science) of children in third grade are positively affected, the effect is, however, not statistically significant. In terms of socio-emotional behaviour (measured by the SDQ filled out by the students' teacher) children seem to fare worse when attending an all-day school with a higher all-day share. The effect is, however, also not statistically significant.

When looking at time use, we see that a 10 percentage point increase in the all-day share is associated with a significant reduction of three minutes in the time spent on homework in second grade. This could be explained by children working on their homework already 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

¹⁸Unfortunately, we do not have information on individual afternoon program attendance by grade but it is very likely that all grades (1 to 4) in primary school are equally targeted by afternoon programs.

¹⁹As described in Section 4.1 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.

is true for 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, yet at the margin of significance. A 10 percentage point increase in the all-day share is estimated to increase the probability of reading 30 minutes per day by three percentage points, the probability of playing PC-games and watching TV for more than two hours per day is reduced by two percentage points. In sum, even if not significant, the pattern of these findings suggests that some substitution of time use takes place when children attend an all-day school. Importantly, the positive effect on teachers' assessment in math and German in fourth grade (see Figure 2) is in line with more time spent on reading.

When evaluating parents' answers on time use (only available when children are in third grade) we find the following. Parents' feeling of spending not enough time with their child increases when the child attends a school with a higher all-day share. The all-day share also has a significant negative effect on the actual hours parents spend with their child on a weekday. On the other hand, hours spent together on weekends are not significantly negatively affected.

The observed children are asked about their satisfaction with their school when they are in third grade. Using this outcome, we find a positive and significant effect of a higher all-day share on school satisfaction. This positive school attitude is also in line with the positive effect of attending an all-day school on teachers' assessments shown in Figure 2.

6 Heterogeneity Analysis

Figure 2 shows the impact of attending an all-day school for the average student. It is, however, very likely that the effect could differ by students' demographics and family backgrounds. In fact, heterogeneous effects are intended as all-day schools are mainly targeted at fostering low performing students' achievement, see Section 2. Furthermore, the counterfactual afternoon program potentially varies by family background.²⁰ Finally, girls and boys may differ in how well they cope with doing homework at their home 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:

$$y_{ism} = \beta' X_{ism} + \gamma Allday_{sm} + \tau \hat{interact}_{ism} + \delta_b + \epsilon_{ism} \quad (3)$$

Where $\hat{interact}_{ism} = Allday_{sm} \times sg_{ism}$ is the product of the all-day share 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,

²⁰For a discussion of the counterfactual situation of formal care in afternoon hours in Germany see Felfe and Zierow 2014 and Felfe and Zierow 2018.

the all-day share 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.4 in the Appendix.

Estimates show that boys benefit more than girls from an increase in the all-day share in the teacher's assessment of their math skills (a 8 ppt increase in their received assessment in contrast to only a 4 ppt increase in case of girls for a 10 ppt increase in the all-day share). This widens the gap between girls and boys with respect to their math performance (as assessed by their teacher) by 28 percent.

For non-native speakers we observe a negative all-day share effect on their achievement measures. They are significantly negatively affected in terms of their test scores in math and German (by 6 ppt and 7 ppt for a 10 ppt increase in the all-day share, respectively), which widens the native vs. non-native gap by 13 and 15 percent. Also with respect to teacher's assessment and high-track recommendation, non-native speakers do not benefit as much from attending an all-day school as the group of native speakers although the estimates for the interaction term are insignificant.

In terms of children living with a single parent, the test score results are not significantly differently affected. The results with respect to teacher assessment and high-track recommendation point towards smaller effects of attending an all-day school in comparison to children living together with both parents. Interestingly, we do not find significant effect heterogeneities by parental educational background.

In sum, our heterogeneity analysis does not provide any evidence for all-day schools causing a reduction in inequality in the German educational system. The results rather suggest that existing student achievement gaps between boys and girls as well as between non-native and native speakers increase with a larger all-day share. A first possible explanation for these results could be subgroup specific selection into afternoon programs offered by all-day schools. As described in Section 2, the afternoon programs are not mandatory, but parents can decide whether they want to enroll their child in the afternoon program offered by an all-day school. Yet, using the limited information provided by the NEPS data on actual enrollment in afternoon programs (only available for the second grade) we do not find differences in enrollment probability by gender. The share of non-native speakers in all-day programs is even larger than their share in the group of non-enrolled students. Thus, subgroup-specific enrollment in an all-day school's afternoon program cannot explain the findings.

Another possible explanation could be the voluntary nature of activities offered to children enrolled in an afternoon program. If students are allowed to sort into voluntary activities in the afternoon hours, the sorting is probably not based on selection-on-gains. If children have the choice, they potentially rather opt for learning activities which they are already good at than

for activities with which they would have difficulties.²¹ In such a scenario, the compensating potential of all-day schools could not unfold.

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 federal all-day school investment project which was set up to promote all-day schools by providing construction subsidies with a volume of over four billion euros in total. We exploit the information on the 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. We do find, however, a positive average effect on children's school performance assessed by their teacher. We also find a positive effect of all-day programs on the teacher's high-track recommendation. Our subgroup analysis reveals that boys benefit more than girls from all-day programs in terms of the assessment they get from their math teacher. Non-native speakers fare worse in terms of their test scores. In sum, we find evidence that all-day school programs benefit children in terms of their school grades and their important transition into the next school track. We cannot, however, provide any proof that the all-day programs contribute to decreasing inequality in the school system.

The positive effects of the all-day share on the teacher's assessments of students are definitely an important finding. Those subjective assessments are known to be associated with non-cognitive skills and motivational dimensions. Both are important concepts and linked to later life outcomes. Increasing them would be a notable success of all-day programs. However, the effects on test scores are found to be insignificant and there is no indication for any reduction in inequality of educational achievements between subgroups. A potential reason could be the voluntary nature of the afternoon activities in all-day schools if children's sorting into these is not based on selection on (cognitive) gains. Mandatory afternoon programs may be better suited to unfold the potential of all-day schools to decrease educational inequality.

References

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

²¹Unfortunately, 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.

- 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 Jackowitz, and Gary Painter (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 and Marie Paul (2017). “The Effects of After-School Programs on Maternal Employment”. In: *SOEPpapers on Multidisciplinary Panel Data Research* 905.
- 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.
- Goodman, Robert (1997). “The Strength and Difficulties Questionnaire: A Research Note”. In: *Journal of Child Psychology and Psychiatry* 38.
- 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.
- Klemm, Klaus and Dirk Zorn, eds. (2017). *Gute Ganztagschule für alle – Kosten für den Ausbau eines qualitativvollen Ganztagschulsystems in Deutschland bis 2030*. 1., überarb. Aufl. Gütersloh: Bertelsmann Stiftung.
- 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 Ganztagschulen als Konsequenz aus PISA? Bildungspolitische Diskurse und Entwicklungen in den Jahren 2000 bis 2003”. In: *Ganztagschule 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 Ganztagschule?: 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.
- 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.
- 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.
- Steinmann, Isa, Rolf Strietholt, and Daniel Caro (2018). “Participation in extracurricular activities and student achievement: Evidence from German all-day schools. School Effectiveness and School Improvement”. In: *School Effectiveness and School Improvement* 44, pp. 1–22.

A Appendix

Table A.1: Control regression

	(1)	
	investment	
county level		
apprenticeship-applicants ratio	-0.822	(0.618)
household income	0.0155	(0.0176)
share of school leavers without certificate	-0.764	(1.289)
share basic high school certificate	-0.258	(0.555)
share university entrance certificate	0.689	(0.572)
municipality level		
labor market participation rate	-0.206	(0.544)
share of children below age 6	-4.742	(3.939)
outward migration	-0.0926	(0.255)
inward migration	0.0883	(0.254)
rate of birth	2.267	(1.889)
Constant	138.9	(72.41)
N	4461	
F	1.468	

Standard errors in parentheses

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

The dependent variable is the instrumental variable, the investments of the IZBB-program (as vingtile median), see Sections 3 and 4.1.

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

The variables on county level are the 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 variables 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		Teacher assessment		
	(1) maths	(2) German	(3) recommendation high track	(4) maths	(5) German
share of students in all-day program	.375 (.434)	.291 (.46)	.414* (.235)	.595* (.342)	.578* (.332)
female	-.125*** (.0283)	.216*** (.0249)	.0261* (.0159)	-.13*** (.0188)	.1*** (.017)
non-native speaker	-.098*** (.0367)	-.0351 (.0329)	-.00556 (.0244)	-.0187 (.0236)	-.0402* (.0212)
age in month	-.0129*** (.00336)	-.00959*** (.00319)	-.00779*** (.002)	-.0043** (.00202)	-.00208 (.00187)
parent, college degree	.429*** (.0321)	.428*** (.0288)	.205*** (.0178)	.163*** (.019)	.19*** (.0181)
single parent household	-.0987** (.0403)	-.0999** (.0421)	-.0891*** (.032)	-.108*** (.0303)	-.113*** (.0255)
sibling in household	.0226 (.0348)	-.0448 (.0343)	-.0258 (.0227)	.00996 (.0239)	-.021 (.0243)
private school	-.0708 (.133)	.0494 (.104)	-.0458 (.0584)	-.0238 (.0832)	-.0461 (.0962)
low ses school	-.0422 (.0443)	-.0701* (.0412)	-.0467 (.0284)	-.0094 (.0316)	-.0262 (.0284)
share of migrant background $\geq 1/3$	-.139** (.0575)	-.0911 (.0624)	.0114 (.0326)	-.0568 (.0482)	-.0299 (.0448)
male teacher	-.0154 (.0585)	-.0373 (.0579)	.0812** (.0354)	.0274 (.0391)	-.00814 (.0413)
teacher younger than 35	-.0908** (.0459)	-.105** (.0515)	-.0408 (.0307)	-.0338 (.0314)	-.0578* (.0305)
full-time teacher	-.0259 (.0455)	.0343 (.0445)	.0111 (.0266)	-.00353 (.0301)	-.0135 (.0319)
GDP p.c., county \geq med.	-.00719 (.052)	-.0706 (.0497)	-.0698** (.0353)	-.0504 (.0359)	-.0424 (.0362)
female employment rate, munc.	.00329 (.00529)	.00359 (.00512)	.000113 (.00322)	.00252 (.00397)	.00345 (.00357)
inhabitans, munc	-.00365 (.0198)	.00489 (.0199)	.000954 (.00994)	-.0152 (.0144)	-.00258 (.0129)
fiscal capicity, munc.	.00441 (.0104)	.00693 (.0103)	.00722 (.0067)	.000325 (.00766)	.001 (.00758)
Constant	-.138 (.388)	-.397 (.393)	.29 (.232)	-.0159 (.344)	-.34 (.316)
FSt	14.5	14	19	10.4	10.7
N	5263	5380	3422	3758	3696

Standard errors in parentheses

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

The test scores are standardized to have mean zero and a standard deviation of one. The teacher assesment are dummies for being recommended to high track in high school and having (very) good skills in math and German.

The share of students in an all-day program ranks from 0 and 1.

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

Table A.3: Sensitivity analysis

	Test scores		Teacher assesment		
	(1) maths	(2) German	(3) recommendation	(4) high track	(5) maths German
estimates for the all-day share:					
baseline regression (see table A.2)	.375	.291	.414*	.595*	.578*
	(.434)	(.46)	(.235)	(.342)	(.332)
FSt	14.5	14	19	10.4	10.7
in a regression with additional controls for:					
additional classes in main subjects	.447	.316	.438*	.616*	.63*
	(.478)	(.502)	(.254)	(.373)	(.369)
FSt	12.4	11.8	16.1	8.7	8.72
high share of qualified employees	.381	.291	.434*	.612*	.588*
	(.448)	(.474)	(.25)	(.344)	(.342)
FSt	16.2	15.5	20	11.4	11.5
support for students at risk	.381	.3	.41*	.609*	.602*
	(.438)	(.456)	(.231)	(.352)	(.348)
FSt	13.5	13	18.7	9.74	9.87
all three variables above	.439	.306	.438*	.624*	.64*
	(.478)	(.493)	(.255)	(.37)	(.38)
FSt	13.8	13.1	18	9.51	9.33
N	5263	5380	3422	3758	3696

Standard errors in parentheses

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

The test scores are standardized to have mean zero and a standard deviation of one. The teacher assesment are dummies for being recommended to high track in high school and having (very) good skills in math and German.

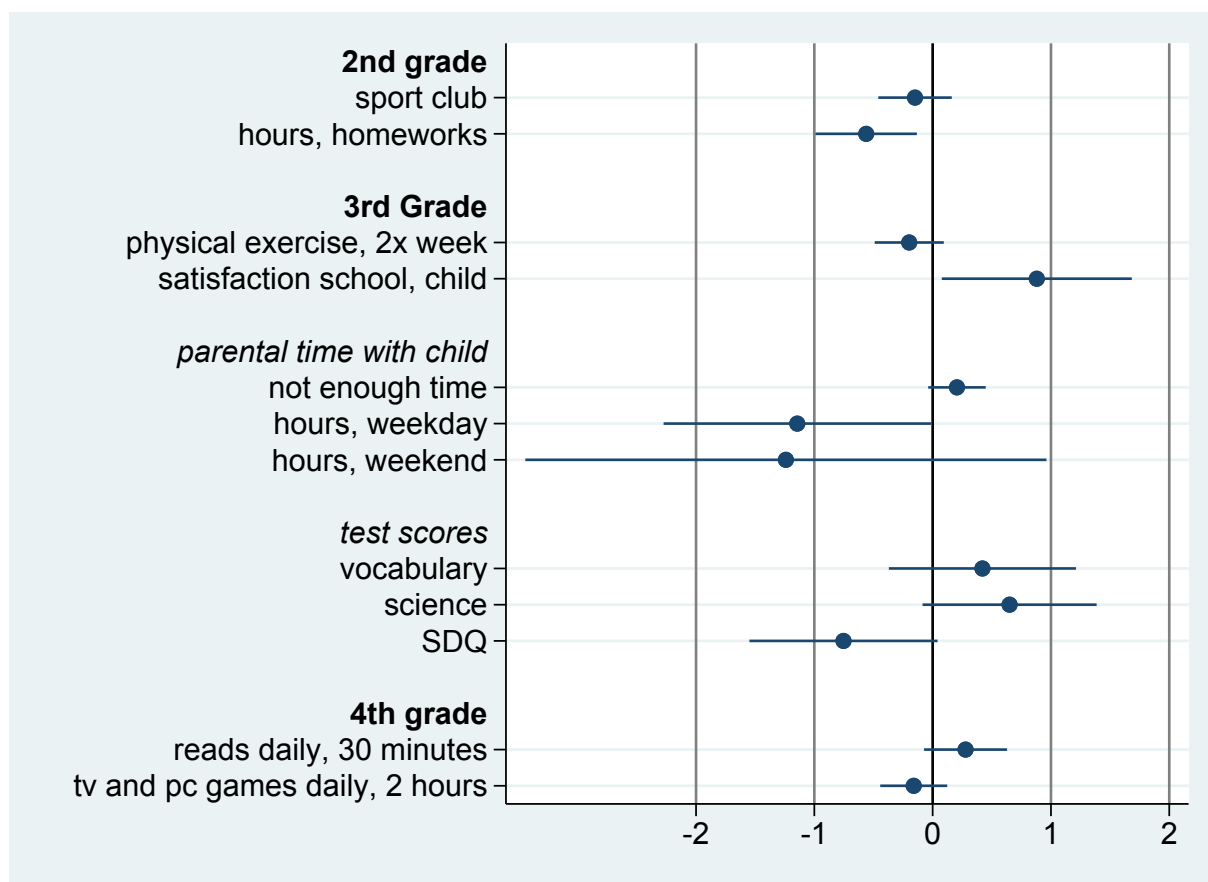
The table shows the estimates for the share of all-day students in the baseline, in a regression with the respective additional control and in a regression controlling for all three auf those variables.

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 unqualified. 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 Table A.2 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.1: The impact of all-day school programs on further outcomes



Notes: Estimates for the all-day share 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 and having a high general satisfaction with the school. Both answers are given by the child. A dummy for not spending enough time with the child and the time spent with the child in hours on weekdays and on weekends. All of these three answered by the parents. And test scores for vocabulary, science and the social behavior as SDQ score (Goodman 1997) evaluated by the teacher, all three standardized.

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 Table A.2 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.4: Heterogeneity analysis

	Test scores		Teacher assessment		
	(1) maths	(2) German	(3) recommendation high track	(4) maths	(5) German
Interaction with female					
share of students in all-day program	.543 (.423)	.345 (.433)	.34 (.227)	.792** (.387)	.707** (.354)
interaction term	-.387 (.355)	-.125 (.306)	.175 (.169)	-.371* (.214)	-.234 (.199)
KPrkSt	6.85	6.65	8.99	5.21	5.41
Interaction with indicator for non-native speakers					
share of students in all-day program	.476 (.449)	.397 (.473)	.428* (.237)	.643* (.35)	.592* (.335)
interaction term	-.64* (.374)	-.708** (.302)	-.107 (.243)	-.493 (.305)	-.124 (.215)
KPrkSt	7.27	7	9.34	5.04	5.28
Interaction with indicator for parents with university diploma					
share of students in all-day program	.28 (.574)	.116 (.533)	.466 (.318)	.625 (.415)	.475 (.411)
interaction term	.143 (.358)	.269 (.331)	-.0753 (.2)	-.0404 (.225)	.138 (.219)
KPrkSt	5.94	5.97	8.16	4.46	4.51
Interaction with indicator for single-parent household					
share of students in all-day program	.327 (.419)	.3 (.465)	.446* (.248)	.581* (.349)	.582* (.336)
interaction term	.478 (.526)	-.102 (.447)	-.303 (.252)	.0895 (.275)	-.03 (.218)
KPrkSt	7.29	7.06	9.48	5.14	5.32
N	5263	5380	3422	3758	3696

Standard errors in parentheses

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

The test scores are standardized to have mean zero and a standard deviation of one. The teacher assessment are dummies for being recommended to high track in high school and having (very) good skills in math and German.

The interaction term is constructed as product of the all-day share and the respective subgroup-identifier. State fixed-effects and the full set of control variables shown in Table A.2 on individual, teacher, school and municipality level controls and controls for missing in those variables are included. Standard errors are clustered on school level.