Routledge
Taylor & Francis Group

School Effectiveness and School Improvement
An International Journal of Research, Policy and Practice

ISSN: 0924-3453 (Print) 1744-5124 (Online) Journal homepage: http://www.tandfonline.com/loi/nses20

Participation in extracurricular activities and student achievement: evidence from German all-day schools

Isa Steinmann, Rolf Strietholt & Daniel Caro

To cite this article: Isa Steinmann, Rolf Strietholt & Daniel Caro (2018): Participation in extracurricular activities and student achievement: evidence from German all-day schools, School Effectiveness and School Improvement, DOI: 10.1080/09243453.2018.1540435

To link to this article: https://doi.org/10.1080/09243453.2018.1540435

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT
https://doi.org/10.1080/09243453.2018.1540435

Routledge
Taylor & Francis Group

ARTICLE

Participation in extracurricular activities and student achievement: evidence from German all-day schools

Isa Steinmann [a], Rolf Strietholt[a] and Daniel Caro[b]

[a]Center for Research on Education and School Development (IFS), TU Dortmund University, Dortmund, Germany; [b]Department of Education, University of Oxford, Oxford, UK

ABSTRACT

The present study investigates the effects of participation in academic extracurricular activities on student achievement in mathematics and reading in the course of secondary schooling. We use 2 samples of nonmandatory all-day schools that are part of the National Educational Panel Study (NEPS): N = 1,131 fifth graders at 43 schools were followed until Grade 7 and N = 1,545 seventh graders at 64 schools until Grade 9. We compare students participating in all-day schools' extracurricular activities homework support, remedial education, and subject-specific programs to nonparticipating students. After controlling for prior achievement and further student background variables as well as estimating school fixed effects, we find no effects of participation in extracurricular activities on student achievement for both samples and both outcomes. The study contributes to the evidence base on the effects of extracurricular activities. We discuss the findings in light of the international discourse on effective extracurricular activities.

ARTICLE HISTORY
Received 26 March 2018
Accepted 22 October 2018

KEYWORDS
All-day school; extracurricular activities; mathematics achievement; reading achievement; secondary schools

Introduction

Much learning takes place outside of regular school lessons. At the age of 15, for example, students in Organisation for Economic Co-operation and Development (OECD) countries spend an average of 18 hr per week learning the language of instruction, mathematics, and science and spend 62% of this time in regular school lessons, 26% studying individually, and 12% in so-called out-of-school-time lessons (OECD, 2011). The out-of-school-time lessons encompass activities taking place at school (e.g., all-day schooling), home (e.g., private tutoring), or elsewhere (e.g., summer schools). School-based extracurricular activities are of particular interest for policymaking and educational research as they are more amenable to regulation by reforms than private study or tutoring (cf. Kuger, 2016).

Two strands of arguments for extending extracurricular activities in schools dominate the public and scientific discourses. From a social policy perspective, the additional time children spend at school should improve the employability of both parents. From an educational perspective, the additional time at school may be used to provide extended learning opportunities and individual support or to enrich and diversify the school day in

CONTACT Isa Steinmann isa.steinmann@tu-dortmund.de

order to encourage the academic and psychosocial development of children (cf. Fischer, Theis, & Züchner, 2014; Plantenga & Remery, 2013; Stecher & Maschke, 2013). We focus on the educational perspective in the present study.

While the desire to support children's development is a universal one, the definition, organization, and use of additional time vary across and within countries. For instance, educational systems differ in terms of their programs (e.g., remedial education, sports clubs), the amount of time allocated to them, staff qualifications, staff-to-child ratios, and enrolment rates (OECD, 2015; Plantenga & Remery, 2013). This international divergence in extended education characteristics needs to be discussed in the context of theoretical effectiveness models. As different models emphasize, extracurricular programs need to meet quality criteria to affect student outcomes like school achievement (cf. Fischer & Klieme, 2013; Miller & Truong, 2009; Vandell, Reisner, & Pierce, 2007). For example, the programs should be designed to foster skills, linked to regular school lessons, and supervised by qualified staff. Furthermore, even if a school provides high-quality programs, development effects can be expected only for those students who actually participate over a certain timespan. If attendance is voluntary, programs need to be attractive to students so that they decide to attend. We can therefore assume that variations between countries in the use of additional time impact the effectiveness of extended education. One can thus expect to find internationally differing effects on outcomes like student achievement: While attending extended learning programs might promote learning in some educational systems, this might not be the case in others.

Extended education and student achievement: international variation in programs and their effectiveness

The term "extended education" refers to programs that differ in their scope and aims. For this reason, it is important to acknowledge such differences when reviewing the previous research on different extended education programs. We limited our literature review to research on student test scores and did not include studies that considered, for example, school grades or motivational outcomes. In the US, so-called after-school programs are rather common. Such programs are mostly organized independently of the schools, operate after regular school finishes, and provide adult supervision for participating students. They differ in terms of their opening hours, facilities (e.g., dedicated facilities vs. community centers), the types of activities (e.g., childcare vs. academic enrichment), and the qualifications of the supervising staff (e.g., trained vs. layperson) (e.g., Scott-Little, Hamann, & Jurs, 2002). Meta-analyses and literature reviews show positive mean effects of after-school programs on student achievement in the US. The primary studies discussed in these reviews are both small-scale interventions and evaluations of interstate initiatives of various program types. Durlak, Weissberg, and Pachan (2010), for example, identified a positive overall effect of after-school programs on student outcomes. Specifically, programs with certain quality criteria had a positive impact on student learning, while there were neutral effects in the absence of such criteria: After-school programs needed to include sequenced learning steps and active student work on dedicated material, be focused in terms of allocated time and attention, and have explicit learning objectives in order to affect student outcomes. The relevance of after-school programs' quality also becomes apparent when reviewing further meta-

analyses and literature reviews. Generally, higher effect sizes were associated with the specificity of the programs' goals, like the aim to promote learning (Apsler, 2009; Cooper, Charlton, Valentine, Muhlenbruck, & Borman, 2000; Lauer et al., 2006), a close link between the programs' contents and regular school curricula (Fashola, 1998; Scott-Little et al., 2002), staff qualifications, and a low teacher–student ratio (Cooper et al., 2000; Fashola, 1998; Feldman & Matjasko, 2005; Scott-Little et al., 2002). Primary school programs were somewhat more effective than programs for older students (Cooper et al., 2000; Scott-Little et al., 2002). However, the positive findings could not be replicated in low-quality programs with opposite characteristics (Apsler, 2009; Durlak et al., 2010; Roth, Malone, & Brooks-Gunn, 2010). Furthermore, researchers identified favorable characteristics of student participation. Students who attended more frequently and consistently (Apsler, 2009; Feldman & Matjasko, 2005), as well as at-risk students, benefited more than others (Patall, Cooper, & Allen, 2010; Scott-Little et al., 2002). Studies on further extracurricular activities like summer schools (which take place during the summer holidays) basically replicate the findings on after-school programs (Lauer et al., 2006).

Compared to the rich body of research on extended education in the US, the number of studies on this topic published in other countries is much smaller. In the following, we provide evaluation examples ranging from small-scale local interventions to large-scale education extensions in entire school systems. We will stress the characteristics of the different investigated programs to illustrate cross-national differences. In Ireland, a small-scale randomized trial found that attending an intervention that aimed to promote reading had literacy-increasing effects over the first school year. This 36-week-long after-school training program was tailored to students who have trouble learning to read in disadvantaged schools and comprised three 90-min sessions per week (Biggart, Kerr, O'Hare, & Connolly, 2013). By contrast, another small-scale randomized field experiment with elementary school students in the Netherlands found no significant effects of attending a different extended day program on mathematics and language achievements. On average, this pilot program comprised 2 additional hours of Dutch instruction, 2 hr of mathematics, and a 1-hr excursion per week. The authors suggest that the intervention period of 11 weeks might have been too short to impact academic achievement (Meyer & Van Klaveren, 2013). In contrast to the after-school programs in the US, which are long term and widely available, the two programs in Ireland and the Netherlands were small pilot interventions that were implemented outside of regular classes but in the school context.

In Italy, a difference-in-difference study found positive effects of a larger scale funding project on mathematics (but not on language) achievement over 1 year in lower secondary schools. The funding enabled more than 200 low-achieving schools to provide additional instruction in language and/or mathematics outside of regular school hours for some classes. All students were required to attend these additional subject-specific activities, which were conducted by specifically trained teachers. Further analyses suggest that the effects were more pronounced for the least advantaged students (Battistin & Meroni, 2016). This project was – similar to the Irish and Dutch studies – implemented in the school context but comprised additional regular classes instead of specific interventions.

In Chile, small mathematics- and language-promoting effects in 10th grade were found over 2 years in a difference-in-difference study on a nation-wide extended education reform. The school days were switched from two shifts – some students had classes in the morning, others in the afternoon – to a full school day for all public schools. In contrast to the previously mentioned European studies, this intervention entailed reforming the whole school system and had very heterogeneous implications for how the additional time was used. Unlike after-schooling in the US, it was closely connected to the schools. Some of the Chilean schools used the additional time for extended regular classes or extracurricular activities that were more or less aligned with learning. The difference-in-difference study compared students who studied under all-day conditions with students who studied under half-day conditions, that is, whose schools were still on the (government-determined) waiting list to implement the reform. Further analyses showed that high-achieving students at rural schools especially benefited from the extended day (Bellei, 2009).

In the German-speaking part of Switzerland, the primary school system comprises both all-day schools and half-day schools. While half-day schools finish before lunch, all-day schools provide lunch and additional supervised time in the afternoon, which is typically used for homework supervision and recreational activities. Participation in these activities is usually voluntary. Students who intensively attended the extracurricular activities at their all-day schools were found to have small mathematics and reading achievement advantages compared to their nonparticipating peers over the first 2 school years. However, these findings need to be interpreted with caution because the treatment groups of the studies were very small (less than 40 students) and included students from more privileged backgrounds (Schüpbach, 2014, 2015).

Similarly, there are all-day schools in addition to traditional half-day schools in Germany. As a result of a massive state funding program that promoted the extension of all-day schools in the beginning of this century, the nation-wide share of all-day schools increased from 16% in 2002 to 65% in 2015 (Kultusministerkonferenz [KMK], 2008, 2017). All-day schools are schools that provide lunch and an all-day program of at least seven hours on at least three days per week. At about half of the all-day schools, participation is nonmandatory: at the other half, it is obligatory either for certain groups of students or for all students. These are called nonmandatory, partially, and fully mandatory all-day schools, respectively (KMK, 2008). Because of the autonomy that German (and Swiss) schools have in designing and organizing their all-day activities, the term all-day schooling refers to programs that are characterized by a low level of standardization in comparison to the previously mentioned programs in other countries. For example, the all-day programs are not necessarily intended to foster learning but also incorporate leisure activities (the consortium of the Study on the Development of All-Day Schools [StEG-Konsortium], 2015, 2016).

In summary, the German state of research is less promising than the international one. In the following, we review well-designed studies aiming to establish causal inference of all-day schooling effects on student achievement. So far, no experiment has been conducted in this field. A study comparing half-day and all-day schools found no significant differences in science, reading, and mathematics achievement. The authors used cross-sectional representative large-scale data from primary and secondary school samples from Germany and applied a matching procedure to balance systematic

differences between half- and all-day schools. Retrospective information on early literacy skills and teacher ratings were used as proxies for prior achievement to mimic a longitudinal design. The analyses further showed that the socioeconomic and immigrant-background-related gaps in student achievement did not differ between half-day and all-day schools (Strietholt, Manitius, Berkemeyer, & Bos, 2015).

Only three longitudinal studies with test measures of prior achievement evaluated the student achievement effects of attending extracurricular activities at all-day schools instead of the aggregated effects on the school level, like Strietholt and colleagues' (2015) study. Bellin and Tamke (2010) found that first-grade students who in general attended the all-day program at their nonmandatory all-day schools attained significantly higher reading scores in second grade after controlling for prior achievement in first grade. However, the effect between the two measurement points was small and no longer significantly different from zero when also controlling for cognitive ability in the first grade. Although these findings are informative, it is important to acknowledge that the authors did not distinguish between domain-specific learning and other activities in the afternoon. In addition, the study was confined to one federal state.

In a second and larger longitudinal study with two measurement points, researchers found that neither participating in reading nor in media programs (e.g., school newspaper or drama club) affected students' reading development between the beginning and middle of the fifth grade (Fischer, Sauerwein, Theis, & Wolgast, 2016). In further analyses, participating students were differentiated according to the voluntary or involuntary nature of their participation: Those who participated voluntarily in reading programs showed achievement advantages over involuntary participants. For media programs, this effect was not significantly different from zero.

In a third longitudinal study with two measurement points, participation in science-related extracurricular activities in all-day schools at the end of the third grade was found to have no significant effect on science achievement at the end of the fourth grade in 10 federal states. Researchers did find positive effects on science achievement when considering participation in programs in the middle or end of the fourth grade. Although these positive effects were significant, the effect sizes were extremely small. The authors therefore concluded that the study could not substantiate a general participation effect. They also tested whether broad participation (i.e., the overall number of attended science programs), continuous participation (i.e., the number of school terms), or the perceived quality of the extracurricular activities impacted achievement but found no evidence to support these hypotheses (Lossen, Tillmann, Holtappels, Rollett, & Hannemann, 2016).

Before summarizing the research on achievement effects, it seems worth remembering that a considerable degree of freedom constitutes German all-day schooling. Overall, however, previous studies show no achievement effect of all-day versus half-day schools or of participation in extracurricular activities at all-day schools in Germany. This also holds for studies that focus explicitly on the effects of domain-specific all-day learning activities on achievement in the respective domain (Fischer et al., 2016; Lossen et al., 2016). Additionally, the longitudinal studies strengthen the assumption that attending after-school activities is subject to selection mechanisms that can bias effect estimations. In the first study, participating students had cognitive and reading advantages over nonparticipating students, and for most of them German was their first language. Controlling for

prior achievement in the analyses did not remove all confounding through unobserved covariates (Bellin & Tamke, 2010). Reading program participants were found to be low-performing readers on average, and media programs were attended by students with higher reading learning goals and higher reading self-concepts on average (Fischer et al., 2016). Attendees of science-related programs tended to be male, have an immigrant background, and a low socioeconomic status (Lossen et al., 2016). Additionally, schools' academic programs differed between school types and were attended by less than 20% of students on average (Fischer et al., 2016; Lossen et al., 2016).

However, the state of research on the effects of all-day schooling on student achievement is scarce and not comprehensive. The studies only look at students from the first to the fifth grades, the comparison intervals are only between a half and 1 year, and they only cover reading- and science-related programs and outcome measures.

The current study

The overview of the effectiveness research reveals considerable differences in how extended education programs have been defined, organized, and studied internationally. These differences reflect a natural limitation of research on extended education across countries. In light of this limitation, we aimed to test whether participation in academic extracurricular activities at German secondary all-day schools affected student achievement in reading and mathematics. Specifically, we focused on a representative sample of nonmandatory all-day schools offering homework support, remedial education, and subject-specific programs where some students did and did not participate in these activities. We focused on this subset of schools in order to exploit the within-schools variation in the explanatory variables and thus circumvent bias from school-level confounding factors.

We separately investigated the following three research questions:

(1) Did students who participated in homework support have achievement advantages over nonparticipating classmates?
(2) Did students who participated in remedial education have achievement advantages over nonparticipants?
(3) Did students who attended subject-specific programs have achievement advantages over those who did not attend such programs?

We studied developments in the core learning areas of reading and mathematics achievement. Following the discussed literature and political expectations, we hypothesized we would find mathematics and reading achievement advantages among students who participated in (a) homework support, (b) remedial education, and (c) subject-specific programs in comparison to nonparticipants.

Method and analyses

Data

This paper uses representative data for Germany on secondary schools from the National Educational Panel Study (NEPS; Blossfeld, Roßbach, & von Maurice, 2011).[1] As we aimed

to investigate the effects of participating in extracurricular activities, we strove to compare students participating in these activities to nonparticipating classmates (see analysis plan section below) and thus focused on nonmandatory all-day schools. We excluded data from special-needs schools because differences between the assessment materials precluded comparisons with the regular school sample. Due to these restrictions, our analyses are limited to regular secondary schools with nonmandatory all-day schooling, which applies to about one out of three secondary schools (KMK, 2017). Additionally, we aimed to investigate achievement developments and therefore only included students who stayed at the same schools between measurement points.

We considered two samples: In the first sample of N = 1,131 students at N = 43 nonmandatory all-day schools, the students were tested in Grades 5 (2010–2011) and 7 (2012–2013). The average age in Grade 5 was M = 10.88 years (SD = 0.53), and 48.3% of the students were girls. The average number of students per school was M = 26.30 (SD = 12.07). The schools were from 10 federal states, and 23 were *Gymnasien* (the school type with the strictest ability selection in the tracked German school system). The second sample comprised data from N = 1,545 students at N = 64 schools who were tested in Grades 7 (2012–2013) and 9 (2014–2015). The average age was M = 12.91 years in Grade 7 (SD = 0.53), and 51.9% of the students were girls. The average number of students per school was M = 24.14 (SD = 9.70). The schools in this second sample were located in 13 federal states, and 32 were *Gymnasien*.

Although there was an overlap across the two samples, we analyzed the data separately for three reasons. First, due to panel attrition, an additional student sample was added to the initial NEPS sample in 2012–2013 for which no prior information was available. Second, only a few students attended specific extracurricular activities consistently between Grades 5 and 9. Third, due to the ongoing reforms many schools changed from half-day to all-day or from nonmandatory to mandatory schooling and vice versa during the investigated timespan. In order to clearly define our treatment, such schools were excluded from the analyses. Treating the two samples separately, however, circumvented some issues related to the treatment status and minimized the number of excluded cases.

Outcomes

The outcome variables were mathematics and reading achievement scores. The paper-pencil tests in mathematics and reading were administered in Grades 5, 7, and 9. The scores were comparable and linked across assessment cycles to make it possible to study change in the respective domains (Leibniz Institute for Educational Trajectories [LIfBi], 2018a, 2018b, 2018c). We used weighted maximum likelihood estimates (WLE) of student abilities based on item response theory models (Pohl & Carstensen, 2012). In Sample 5–7, the mean WLE posttest scores in Grade 7 were M = 0.99 in mathematics (SD = 1.18, 5.7% missing) and M = 0.87 in reading (SD = 1.31, 5.8% missing). In Sample 7–9, the mean achievement scores in Grade 9 were M = 1.71 in mathematics (SD = 1.24, 8.2% missing) and M = 1.29 in reading (SD = 1.22, 16.0% missing).

The mathematics test assessed student achievement in abstract thinking and mathematical problem solving in the four content areas quantity (e.g., percentage calculation), space and shape (e.g., geometric figures), change and relationship (e.g., algebraic

curves), as well as data and chance (e.g., probability calculation). The expected a posteriori/plausible value (EAP/PV) reliability of the test was 0.802 and was determined on the basis of all fifth graders taking the NEPS test (Duchhardt & Gerdes, 2012). The reading test assessed how students derived information from continuous texts, drew conclusions, reflected, and assessed contents. The reading material covered informational, commenting, literary, instructional, and advertising texts. The EAP/PV reliability of the test was 0.811 and was also determined based on the performance of all tested fifth graders (Pohl, Haberkorn, Hardt, & Wiegand, 2012).

Treatments

The explanatory variables were different forms of participation in three extracurricular activities: (a) "homework support/supervision", (b) "remedial groups/remedial education", and (c) "subject-specific learning offers (e.g., extra or enhancement courses in math or German)". The binary variables were 0 if a student did not attend the respective activity and 1 if he or she did attend. The variables were administered in Grades 6 and 8. In Sample 5–7, 17.2% of students took part in homework support (9.4% missing), 15.5% in remedial education (9.8% missing), and 7.4% in subject-specific programs (10.8% missing). The participation rates were lower in Sample 7–9: 8.3% in homework support (10.8% missing), 8.0% in remedial education (11.0% missing), and 6.3% in subject-specific programs (11.4% missing). In Sample 5–7, only 5.0% attended two and 1.2% all three activities. In Sample 7–9, 4.4% attended two and 1.6% three activities. In nonmandatory all-day schools, participation in extracurricular activities must be declared in advance and is binding for at least half a school year (KMK, 2008), which therefore reflects the minimum treatment dosage.

Covariates

Participation in extracurricular activities at nonmandatory all-day schools is not randomized. The literature shows that the characteristics of extracurricular programs vary between federal states and individual schools, for instance, in terms of participation rates (Fischer & Klieme, 2013; Prein, Rauschenbach, & Züchner, 2009; StEG-Konsortium, 2015; Steiner, 2011). Previous research also documents systematic differences at the individual level. Students attending extracurricular activities differ from their nonparticipating classmates in terms of cognitive and family background measures. The degree and direction of these differences depend on the program type (Bellin & Tamke, 2010; Fischer & Klieme, 2013; Fischer et al., 2016; Lossen et al., 2016; Prein et al., 2009; Schüpbach, 2014, 2015; Steiner, 2011). On the basis of the literature, we determined a series of covariates reflecting selection mechanisms at student level. These include prior cognitive and noncognitive measures (tests, grades, school satisfaction) and family background (social and employment status). We also added measures of alternative parental learning support that might interrelate with enrollment in extracurricular activities (homework help by parents, private tutors) (see Appendix 1).

In order to illustrate selection mechanisms, Table 1 compares the covariate statistics of both participating and nonparticipating students in Sample 5–7 with regard to all three treatments. The findings suggest that these treatment and control groups had

Table 1. Treatment and control groups' comparison statistics of covariates in Sample 5–7.

	Homework support			Remedial education			Subject-specific programs		
	M(*SD*) or % in categ. treated	*M*(*SD*) or % in categ. controls	*p* value group difference	*M*(*SD*) or % in categ. treated	*M*(*SD*) or % in categ. controls	*p* value group difference	*M*(*SD*) or % in categ. treated	*M*(*SD*) or % in categ. controls	*p* value group difference
Mathematics pretest	−0.33	0.27	0.00	−0.17	0.23	0.00	−0.02	0.19	0.12
	(1.14)	(1.08)		(1.01)	(1.12)		(1.04)	(1.12)	
Reading pretest	−0.22	0.23	0.00	−0.30	0.24	0.00	−0.08	0.17	0.08
	(1.32)	(1.16)		(1.11)	(1.20)		(1.14)	(1.20)	
Latest school grade mathematics	3.61	3.91	0.00	3.67	3.89	0.00	3.75	3.87	0.36
	(0.90)	(0.83)		(0.78)	(0.86)		(0.82)	(0.85)	
Latest school grade German	3.48	3.89	0.00	3.55	3.87	0.00	3.64	3.83	0.09
	(0.86)	(0.78)		(0.87)	(0.79)		(0.92)	(0.80)	
School satisfaction	7.51	7.86	0.13	7.21	7.90	0.00	7.72	7.80	0.78
	(2.57)	(2.37)		(2.59)	(2.36)		(2.47)	(2.40)	
Books at home	2.75	3.12	0.00	2.91	3.08	0.19	2.93	3.06	0.49
	(1.37)	(1.30)		(1.26)	(1.33)		(1.30)	(1.32)	
Parent employment	80.3%	71.1%	0.15	67.9%	73.6%	0.18	53.5%	74.2%	0.00
	yes	yes		yes	yes		yes	yes	
Homework help parents	2.85	2.65	0.03	2.57	2.70	0.19	2.67	2.68	0.94
	(1.03)	(0.98)		(0.94)	(1.00)		(1.03)	(0.99)	
Homework help tutor	0.52	0.17	0.00	0.49	0.18	0.00	0.48	0.21	0.04
	(1.15)	(0.63)		(1.11)	(0.66)		(1.13)	(0.72)	

Note: Based on weighted imputed data. N = 1,131 students at 43 schools.

heterogeneous characteristics. For example, students who participated in homework support or remedial education had significantly lower pretest scores in mathematics and reading, poorer grades, and more often received homework help from a tutor than nonparticipants. With regard to subject-specific programs, participating and nonparticipating students differed in fewer respects. In Sample 7–9, we found similar differences in student background variables that reflect mechanisms of selection into treatment (see Appendix 2).

Missing values

We imputed missing data that occured at the student level with regard to all outcomes, treatments, and covariates five times by using predictive mean matching. The imputation model used all aforementioned treatments, outcomes, and covariates along with further variables from student, parent, and principal surveys, which are all listed in Appendix 3. The same variables were included in the imputation models of both samples. We included respective sampling weights to replicate the data structure. Dummy variables for school identification were included to account for the nested data structure (cf. Lüdtke, Robitzsch, & Grund, 2017). All further analyses were replicated for the five imputed datasets, and the estimates were combined using Rubin's (1987) rules.

Analysis plan

A standard issue in observational studies is that the treatment may be correlated with other predictors of the outcome. In the present study, we used longitudinal data from a representative sample of German nonmandatory all-day schools to estimate the effects of attending extracurricular activities on student achievement. The empirical issues can be depicted in a simple linear model in which the achievement Y of student i in school s is a function of his or her participation in extracurricular activities (E) and other confounding variables. It is useful to distinguish between confounding variables at the individual (I) and school (S) levels:

$$Y_{is} = b_0 + b_1 E_{is} + b_2 I_{is} + b_3 S_s + e_{is} \quad (1)$$

We were mainly interested in estimating b_1, which is the effect of attending an extracurricular activity on achievement, while holding other predictors of student performance constant. To identify this parameter, the error term e_{is} has to be orthogonal to other explanatory variables and especially to participation in extracurricular activities. We applied two strategies to meet this assumption at school and individual levels.

At the school level, unobserved school composition or school quality measures may be correlated with the provision of extracurricular courses. If, for example, schools operating in deprived areas provide more extracurricular learning opportunities, children from disadvantaged families might participate more frequently in such activities. Hence, the estimation of the effect of extracurricular activities would be biased because e_{is} would be correlated with the explanatory variable. In the same vein, other school measures like educational tracks may be correlated with the supply of extracurricular activities. To avoid bias from any selection mechanisms at a school level, we estimated a fixed effects model to which we added school dummies (μ_s). School fixed effects can be

interpreted as different intercepts for all schools, which is why any variation in outcomes owed to school characteristics is absorbed in these school fixed effects. By implication, we estimated b_1 based on the variation in participation in extracurricular activities within schools (i.e., we compared participating and nonparticipating students from the same school). The fact that this approach avoids school feature bias is reflected in Equation (2), which no longer contains any school-level covariates:

$$Y_{is} = b_0 + b_1 E_{is} + b_2 I_{is} + \mu_s + e_{is} \quad (2)$$

Selection mechanisms within schools remain an issue in the identification of the effects of extracurricular activities. For example, students who are low performing or from disadvantaged families were found to enroll in homework support more often (see Table 1 and Appendix 2). We attempted to address this issue by including extensive controls for prior achievement and family background measures. For this purpose, we used rich test and survey information from achievement tests, student questionnaires, and parent computer-assisted telephone interviews that were conducted prior to the treatment (see Appendix 1). Furthermore, some students attended more than one extracurricular activity at a time, which could have led to an overestimation of the effects of the respective activities. To address this issue, we included all three treatment variables simultaneously in the analysis models.

One tradeoff that arises when using extensive control variables is the increasing risk of an overadjustment bias, where some covariates are correlated with the error term. This leads to an underestimation of the treatment effects (see Castellano, Rabe-Hesketh, & Skrondal, 2014). To address this issue, we introduced the covariates stepwise.[2]

We replicated the analyses for two outcomes (mathematics, reading) and two samples (Sample 5–7, Sample 7–9), resulting in four sets of models. The NEPS target population are students in schools offering lower secondary education in Germany. NEPS applied a stratified two-stage probability-proportional-to-size sampling strategy to sample schools as primary sampling units at the first stage and students at the second stage. We used sampling weights to account for different sampling probabilities as well as initial and wave-specific nonresponse in order to generalize our findings to the underlying student population (for technical details, see Steinhauer & Zinn, 2016).

Results

As a baseline for our main models with school fixed effects and extensive control variables, we regressed the outcomes on participation in the three extracurricular activities without any further controls. These analyses estimated the cross-sectional association between the outcomes and participation in extracurricular activities. In the sample from Grades 5 to 7, students who had attended homework support or remedial education in Grade 6 had lower mathematics and reading achievement scores in Grade 7. Participation in subject-specific programs was not significantly associated with performance levels (see Table 2, columns 1 and 5). In Sample 7–9, only participating in homework support was significantly related to lower mathematics achievement in Grade 9 (see Table 3, columns 1 and 5). However, cross-sectional estimates were affected by selection mechanisms at the school and individual levels and should not be interpreted as causal effects.

Table 2. Treatment effects on mathematics and reading in Sample 5–7.

	Mathematics				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Homework support	−0.40*	0.01	0.03	0.05	−0.39*	−0.07	−0.04	−0.11
	(0.11)	(0.08)	(0.08)	(0.08)	(0.11)	(0.09)	(0.09)	(0.11)
Remedial education	−0.26*	0.00	−0.01	−0.03	−0.38*	−0.09	−0.08	−0.10
	(0.11)	(0.07)	(0.07)	(0.08)	(0.13)	(0.11)	(0.10)	(0.11)
Subject-specific programs	−0.02	0.06	0.06	0.01	−0.35	−0.26	−0.25	−0.29
	(0.15)	(0.11)	(0.11)	(0.11)	(0.18)	(0.14)	(0.14)	(0.15)
Mathematics pretest	–	0.65*	0.56*	0.55*	–	0.28*	0.25*	0.23*
		(0.03)	(0.03)	(0.03)		(0.04)	(0.04)	(0.05)
Reading pretest	–	0.18*	0.17*	0.16*	–	0.49*	0.45*	0.41*
		(0.03)	(0.03)	(0.03)		(0.04)	(0.04)	(0.04)
Latest school grade mathematics	–	–	0.28*	0.24*	–	–	0.03	−0.03
			(0.04)	(0.04)			(0.05)	(0.05)
Latest school grade German	–	–	−0.03	−0.06	–	–	0.21*	0.17*
			(0.05)	(0.05)			(0.05)	(0.05)
School satisfaction	–	–	−0.01	−0.01	–	–	0.00	0.01
			(0.01)	(0.01)			(0.01)	(0.01)
Books at home	–	–	0.05*	0.03	–	–	0.00	0.00
			(0.02)	(0.02)			(0.03)	(0.03)
Parent employment	–	–	0.00	−0.06	–	–	0.06	0.04
			(0.06)	(0.07)			(0.08)	(0.09)
Homework help parents	–	–	−0.02	−0.03	–	–	−0.02	−0.03
			(0.03)	(0.03)			(0.04)	(0.04)
Homework help tutor	–	–	0.05	0.04	–	–	0.06	0.05
			(0.04)	(0.04)			(0.05)	(0.05)
School fixed effects	no	no	no	yes	no	no	no	yes

Note: Based on weighted imputed data; unstandardized *b* values; standard errors in parentheses; *$p < 0.05$; $N = 1{,}131$ students at 43 schools.

In the main analyses, we regressed the outcomes on the treatment variables while controlling stepwise for prior achievement, further student covariates, and school fixed effects. After controlling for prior achievement, the effects of the treatments homework support, remedial education, and subject-specific programs on mathematics and reading achievement were not significantly different from zero (Tables 2 and 3, columns 2 and 6). The results were qualitatively the same after controlling for further student covariates (Tables 2 and 3, columns 3 and 7) and estimating school fixed effects (Tables 2 and 3, columns 4 and 8). These findings further support the assumption that negative associations between outcomes and treatments in models without covariate control and school fixed effect estimation were confounded by selection mechanisms. The main analyses provided no evidence for an effect of participation in extracurricular activities on student achievement.

Discussion

The present study investigates whether attending academic extracurricular activities results in learning benefits for students at nonmandatory all-day schools in Germany. The results suggest that participating in homework support, remedial education, and subject-specific programs has no effects on mathematics and reading achievements in the course of secondary schooling. These results are in line with previous research that showed no effects of attending all-day schools' extracurricular activities on achievement in Germany (Bellin & Tamke, 2010; Fischer et al., 2016; Lossen et al., 2016).

Table 3. Treatment effects on mathematics and reading in Sample 7–9.

	Mathematics				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Homework support	−0.39*	−0.10	−0.11	−0.09	−0.06	0.17	0.16	0.07
	(0.16)	(0.12)	(0.12)	(0.12)	(0.17)	(0.15)	(0.15)	(0.15)
Remedial education	−0.18	0.01	0.09	0.11	−0.12	0.00	0.03	−0.05
	(0.16)	(0.14)	(0.13)	(0.10)	(0.15)	(0.12)	(0.12)	(0.10)
Subject-specific programs	−0.03	0.03	0.04	0.02	−0.12	−0.03	−0.02	0.14
	(0.19)	(0.15)	(0.15)	(0.10)	(0.20)	(0.16)	(0.16)	(0.10)
Mathematics pretest	–	0.58*	0.51*	0.46*	–	0.20*	0.19*	0.12*
		(0.07)	(0.06)	(0.03)		(0.05)	(0.05)	(0.04)
Reading pretest	–	0.14*	0.14*	0.08*	–	0.39*	0.35*	0.30*
		(0.03)	(0.03)	(0.02)		(0.03)	(0.03)	(0.03)
Latest school grade mathematics	–	–	0.15*	0.21*	–	–	−0.03	0.00
			(0.06)	(0.03)			(0.04)	(0.03)
Latest school grade German	–	–	0.00	0.01	–	–	0.15*	0.15*
	–		(0.06)	(0.05)			(0.06)	(0.05)
School satisfaction	–	–	−0.01	−0.01	–	–	−0.02	−0.02
			(0.02)	(0.01)			(0.01)	(0.01)
Books at home	–	–	0.03	0.03	–	–	0.06	0.04
			(0.03)	(0.02)	–	–	(0.03)	(0.02)
Parent employment	–	–	−0.16	−0.08	–	–	0.03	0.00
			(0.21)	(0.07)		–	(0.11)	(0.10)
Homework help parents	–	–	−0.02	−0.01	–	–	−0.03	−0.02
			(0.06)	(0.03)			(0.03)	(0.03)
Homework help tutor	–	–	−0.07	−0.03	–	–	−0.01	0.00
			(0.07)	(0.02)			(0.03)	(0.03)
School fixed effects	no	no	no	yes	no	no	no	yes

Note: Based on weighted imputed data; unstandardized *b* values; standard errors in parentheses; *$p < 0.05$; $N = 1{,}545$ students at 64 schools.

Another strand of findings concerns participation characteristics. We found relatively low rates of attendance in the investigated academic program types. Further, we found selection mechanisms associated with student attendance. Students who participated in homework support and remedial education especially tended to come from a disadvantaged academic and social background. These findings are in line with theoretical and empirical deliberations in the literature, too (Fischer & Klieme, 2013; Fischer et al., 2016; Lossen et al., 2016; Steiner, 2011). We also found that students attending the programs more frequently made use of private tutoring. Considering the remedial nature of the selection mechanisms, this might not be surprising. However, in light of international research in this field, one might conclude that observed treatments may relate to unobserved additional treatments.

With regard to policymaking, the major result of our study is that the investigated forms of all-day schooling are not found to foster student learning and therefore seem insufficiently designed and organized for this particular purpose. International examples show that programs with high-quality characteristics can serve as valuable supports for student learning (e.g., Apsler, 2009; Biggart et al., 2013; Durlak et al., 2010; Lauer et al., 2006). This implies that promising international findings could be replicated when investigating high-quality extracurricular activities. However, they remain rare in Germany. For example, the schools' extracurricular programs are not necessarily connected to the morning classes or supervised by teachers or other highly qualified staff (e.g., Fischer & Klieme, 2013; StEG-Konsortium, 2015, 2016; Steiner, 2011).

The present study has substantive and methodological merits and limitations, as the dataset it used was not specifically designed for investigating the present research questions. A first set of issues relates to the measures of the explanatory variables. We used fairly simple treatment variables, namely, the participation in three extracurricular activities (homework support, remedial education, subject-specific programs). An advantage of NEPS is that it surveys various common activities, including programs with an explicit focus on learning, as investigated in the analyses. At the same time, future research may investigate less common but possibly effective academic activities that were not surveyed in NEPS. In this regard, it may also be worth investigating domain-specific courses because such specific interventions are conceptually more closely related outcomes in the respective domains. However, previous studies on such domain-specific programs found no significant participation effects (Fischer et al., 2016; Lossen et al., 2016). Additionally, even nonacademic treatments might affect student learning indirectly (e.g., participation in sports may increase cognitive capacities). Another limitation relates to how the participation in all-day activities was measured in NEPS. The binary treatments of participation versus nonparticipation were only assessed once (in Grade 6 and 8, respectively) between the achievement tests. Therefore, their minimal dosage was a half year of participation, and we were not able to investigate more continuous attendance.

Apart from the simple treatment variables, it must be acknowledged that we were unable to further investigate possible mediating roles of program quality such as the qualification of supervising staff. One possible explanation for our neutral results is that the quality of extracurricular activities in German all-day schools is too low on average. This explanation is largely consistent with meta-analyses showing that extracurricular activities have no effect on student achievement unless they meet high quality standards (Apsler, 2009; Durlak et al., 2010; Roth et al., 2010). The heterogeneous quality of German all-day schools and of the activities can generally be expected to lead to heterogeneous effects on student outcomes. This relates to general limitations of research with observational data. Given that we find a neutral overall effect, however, this would imply that negative effects in some schools counteract positive effects in others. Following this argument, the recent expansion of all-day schools in Germany apparently did not lead to the development of high-quality extracurricular programs in the afternoon, at least not overall (cf. also StEG-Konsortium, 2016). In the present study, we were however only able to investigate the overall effect of unstandardized treatments that are subsumed under the terms homework support, remedial education, and subject-specific programs. Even students from the same schools might have attended different courses for homework support, for instance. Future research with experimental designs could provide insights into the effects of well-defined extracurricular interventions on closely connected student outcomes.

Furthermore, the scope of our study is limited to specific educational outcomes. Although we considered key measures of student achievement in different grades and the core domains of mathematics and reading, there are other important outcomes such as emotions, attitudes, social and problematic behavior (e.g., drug use), or school dropout (cf. Fischer & Klieme, 2013; Fischer et al., 2014; StEG-Konsortium, 2016; Vandell et al., 2007). For example, previous studies found that all-day schooling positively affected student outcomes like learning goal orientation, academic self-concepts,

motivation, social behavior, and school grades. Typically, such effects were mediated by the programs' quality characteristics and by continuous and intensive student participation (cf. Fischer et al., 2016; Kuhn & Fischer, 2011; Lossen et al., 2016; Sauerwein, Theis, & Fischer, 2016; StEG-Konsortium, 2016). The inconsistent finding on effects of all-day schooling on school grades and test scores in particular deserves closer attention. If one assumes that teacher judgments are not only influenced by students' performance but also by their behavior (e.g., completion of homework, social behavior in class) or motivation, the inconsistency may suggest that all-day schooling affects behavioral and motivational constructs rather than performance.

For a holistic and realistic evaluation of all-day schooling, it seems worthwhile acknowledging noneducational outcomes. The increase in parental labor-market participation is another aim of all-day schools. Some initial studies suggest that the recent extension of all-day schools in Germany has had a positive impact on mothers' labor market involvement (Bauernschuster & Schlotter, 2015; Gambaro, Marcus, & Peter, 2016). This finding reflects that extending all-day schools has led to an increase in the time students spend at school. From an educational perspective, however, a simple extension of time seems to be insufficient to affect students' learning outcomes measured in standardized tests.

Other issues are linked to the identification of effects with nonexperimental data. As a first step, our analytical approach relies on school fixed effects that effectively control for any confounding variables at the school level and by implication also at higher levels. Within schools, we then considered student-level confounders because students (or their parents) select into participation. We sought to minimize the risk of confounded effect estimates by using rich pretest and background data as controls. The most important remaining risk is time-invariant confounding, which is correlated with the treatment and outcome variables but not with the observed covariates.[3] Related to the previous point, the control of pretest and key background measures in linear models may not be effective if the assumption of predictor linearity does not hold. We partly addressed this issue by replicating the analyses using nonparametric propensity score matching in unreported analyses. These analyses qualitatively replicated the findings of our main analyses. However, two limitations must be stressed. First, although we applied different matching algorithms, we were not able to remove all differences between the treatment and control groups. A reasonable explanation for this is that the exact matching on the school makes it difficult to find good matches because the number of classmates is limited within schools. A second limitation could be even more complex selection mechanisms, which would have to be modeled as interaction terms between covariates, for example. Again, sample size sets a natural limitation for investigations with the present data.

A last methodological note relates to the regression-towards-the-mean phenomenon. We observed that participation in extracurricular activities is partly due to a selection based on prior school grades. Such differences may not only result from actual performance differences but also in part from chance. If so, regression towards the mean could mask the positive effects of extracurricular activities on achievement, resulting in a neutral overall effect. Again, as we included not only grades but also various test scores as controls, we think that regression toward the mean did not considerably bias our effect estimates.

However, our results should be interpreted with caution, as it is difficult – even with longitudinal data – to identify the impact of attending extracurricular activities on student achievement. Obviously, it would be easier to identify this effect in carefully implemented small-scale interventions with randomized group allocation. However, the external validity of results is a common problem for experimental approaches, which is why we believe that the present study is nonetheless an important complement to the state of research. We think that our results should contribute to policy discussions on the quality of extracurricular activities when they are implemented on a large scale.

Notes

1. Starting Cohort Grade 5; doi:10.5157/NEPS:SC3:7.0.1. From 2008 to 2013, NEPS data were 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.
2. A further issue of regression adjustments relates to the linearity assumption. To address this issue, we applied nonparametric propensity score matching to adjust for confounding variables (Stuart, 2010). In the matching model, we included all above-mentioned student covariates and used 1:1 nearest-neighbor matching with replacement and calipers. Furthermore, we conducted exact matching on the school identification variable. The matching analyses lead to qualitatively the same results as the regression adjustments.
3. However, controlling for observed covariates also (at least partly) controls for unobserved covariates if they are correlated with the observed ones (Stuart, 2010).

Acknowledgements

We would like to thank Roisin Cronin for copy editing the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Isa Steinmann is a research fellow at the Center for Research on Education and School Development (IFS) in Dortmund. Her main research interests lie in the field of school development – especially the evaluation of all-day schooling in Germany – as well as international comparative research.

Rolf Strietholt is a researcher at the Center for Research on Education and School Development (IFS) in Dortmund. His main research interests comprise educational effectiveness research, heterogeneity, inequality and justice, international comparative research, and empirical social research.

Daniel Caro is a research fellow at the Oxford University Centre for Educational Assessment (OUCEA). His research interests include education inequality, international large-scale student assessments, R programming, mixed models in cross-sectional and longitudinal settings, and causal inference with observational data.

ORCID

Isa Steinmann http://orcid.org/0000-0002-9940-4413

References

Apsler, R. (2009). After-school programs for adolescents: A review of evaluation research. *Adolescence*, *44*(173), 1–19.

Battistin, E., & Meroni, E. C. (2016). Should we increase instruction time in low achieving schools? Evidence from Southern Italy. *Economics of Education Review*, *55*, 39–56. doi:10.1016/j.econedurev.2016.08.003

Bauernschuster, S., & Schlotter, M. (2015). Public child care and mothers' labor supply – Evidence from two quasi-experiments. *Journal of Public Economics*, *123*, 1–16. doi:10.1016/j.jpubeco.2014.12.013

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review*, *28*(5), 629–640. doi:10.1016/j.econedurev.2009.01.008

Bellin, N., & Tamke, F. (2010). Bessere Leistungen durch Teilnahme am offenen Ganztagsbetrieb? [Does full-day school attendance lead to higher achievement?]. *Empirische Pädagogik*, *24*(2), 93–112.

Biggart, A., Kerr, K., O'Hare, L., & Connolly, P. (2013). A randomised control trial evaluation of a literacy after-school programme for struggling beginning readers. *International Journal of Educational Research*, *62*, 129–140. doi:10.1016/j.ijer.2013.07.005

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). Education as a lifelong process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*.

Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Composition, context, and endogeneity in school and teacher comparisons. *Journal of Educational and Behavioral Statistics*, *39*(5), 333–367. doi:10.3102/1076998614547576

Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., & Borman, G. D. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, *65*(1), 1–127.

Duchhardt, D., & Gerdes, A. (2012). *NEPS technical report for mathematics – Scaling results of starting cohort 3 in fifth grade* (NEPS Working Paper No. 19). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American journal of community psychology*, *45*(3–4), 294–309. doi:10.1007/s10464-010-9300-6

Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness* (Report: No. 24). Baltimore, MD: Center for Research on the Education of Students Placed at Risk.

Feldman, A. F., & Matjasko, J. L. (2005). The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions. *Review of Educational Research*, *75*(2), 159–210. doi:10.3102/00346543075002159

Fischer, N., & Klieme, E. (2013). Quality and effectiveness of German all-day schools. Results of the study on the development of all-day schools in Germany. In J. Ecarius, E. Klieme, L. Stecher, & J. Woods (Eds.), *Extended education – An international perspective: Proceedings of the international conference on extracurricular and out-of-school time educational research* (pp. 27–52). Opladen: Barbara Budrich.

Fischer, N., Sauerwein, M. N., Theis, D., & Wolgast, A. (2016). Vom Lesenlernen in der Ganztagsschule: Leisten Ganztagsangebote einen Beitrag zur Leseförderung am Beginn der Sekundarstufe I? [Learning to read in the all-day school: Do all-day programs contribute to reading promotion in the beginning of lower secondary schooling?]. *Zeitschrift für Pädagogik*, *62*(6), 780–796.

Fischer, N., Theis, D., & Züchner, I. (2014). Narrowing the gap? The role of all-day schools in reducing educational inequality in Germany. *International Journal for Research on Extended Education*, *2*(1), 79–96.

Gambaro, L., Marcus, J., & Peter, F. (2016). *School entry, afternoon care and mother's labour supply* (DIW Discussion Paper No. 1622). Retrieved from http://hdl.handle.net/10419/148318

Kuger, S. (2016). Curriculum and learning time in international school achievement studies. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning: An international perspective* (pp. 395–422). Cham: Springer.

Kuhn, H. P., & Fischer, N. (2011). Entwicklung der Schulnoten in der Ganztagsschule – Einflüsse der Ganztagsteilnahme und der Angebotsqualität [Development of school grades in the all-day school – Influences of the all-day participation and the program quality]. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher, & I. Züchner (Eds.), *Ganztagsschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagsschulen (StEG)* (pp. 207–226). Weinheim: Juventa.

Kultusministerkonferenz. (2008). *Allgemein bildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland: Statistik 2002 bis 2006* [Schools of general education in all-day form in the states of the Federal Republic of Germany: Statistics 2002 until 2006]. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_03_04-Allgem-Schulen-Ganztagsform-02-06.pdf

Kultusministerkonferenz. (2017). *Allgemeinbildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland* [Schools of general education in all-day form in the states of the Federal Republic of Germany]. Retrieved from https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/allgemeinbildende-schulen-in-ganztagsform.html

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*(2), 275–313. doi:10.3102/00346543076002275

Leibniz Institute for Educational Trajectories (LIfBi). (2018a). *Information on competence testing: NEPS starting cohort 3 – grade 5: Paths through lower secondary school – Educational pathways of students in grade 5 and higher: Wave 1: grade 5*. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/1-0-0/NEPS_SC3_Competences_W1_en.pdf

Leibniz Institute for Educational Trajectories (LIfBi). (2018b). *Information on competence testing: NEPS starting cohort 3 – grade 5: Paths through lower secondary school – Educational pathways of students in grade 5 and higher: Wave 3: grade 7*. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/3-1-0/NEPS_SC3_Competences_W3_en.pdf

Leibniz Institute for Educational Trajectories (LIfBi). (2018c). *Information on competence testing: NEPS starting cohort 3 – grade 5: Paths through lower secondary school – Educational pathways of students in grade 5 and higher: Wave 6: grade 9*. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/6-0-0/NEPS_SC3_Competences_W6_en.pdf

Lossen, K., Tillmann, K., Holtappels, H. G., Rollett, W., & Hannemann, J. (2016). Entwicklung der naturwissenschaftlichen Kompetenzen und des sachunterrichtsbezogenen Selbstkonzepts bei Schüler/-innen in Ganztagsgrundschulen: Ergebnisse der Längsschnittstudie StEG-P zu Effekten der Schülerteilnahme und der Angebotsqualität [Development of science competencies and the science-related self-concept of pupils in all-day primary schools: Results of the longitudinal study StEG-P on the effects of student participation and program quality]. *Zeitschrift für Pädagogik*, *62*(6), 760–779.

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, *22*(1), 141–165. doi:10.1037/met0000096

Meyer, E., & Van Klaveren, C. (2013). The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands. *Economics of Education Review*, *36*, 1–11. doi:10.1016/j.econedurev.2013.04.002

Miller, B. M., & Truong, K. A. (2009). The role of afterschool and summer in achievement. The untapped power of afterschool and summer to advance student achievement. *Zeitschrift für Pädagogik*, *55* (54th supplement), 124–142.

Organisation for Economic Co-operation and Development. (2011). *PISA quality time for students: Learning in and out of school*. Paris: Author. doi:10.1787/9789264087057-en

Organisation for Economic Co-operation and Development. (2015). *PF4.3: Out-of-school-hours care services*. Retrieved from http://www.oecd.org/els/soc/PF4-3-Out-of-school-hours-care.pdf

Patall, E. A., Cooper, H., & Allen, A. B. (2010). Extending the school day or school year: A systematic review of research (1985–2009). *Review of Educational Research*, *80*(3), 401–436. doi:10.3102/0034654310377086

Plantenga, J., & Remery, C. (2013). *Childcare services for school age children: A comparative review of 33 countries*. Luxembourg: Publications Office of the European Union.

Pohl, S., & Carstensen, C. H. (2012). *NEPS technical report – Scaling the data of the competence tests* (NEPS Working Paper No. 14). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Pohl, S., Haberkorn, K., Hardt, K., & Wiegand, E. (2012). *NEPS technical report for reading – Scaling results of starting cohort 3 in fifth grade* (NEPS Working Paper No. 15). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Prein, G., Rauschenbach, T., & Züchner, I. (2009). Analysen zur Selektivität von offenen Ganztagsschulen [Analyses on the selectivity of nonmandatory all-day schools]. In F. Prüß, S. Kortas, & M. Schöpa (Eds.), *Die Ganztagsschule: vonder Theorie zur Praxis: Anforderungen und Perspektiven für Erziehungswissenschaften und Schulentwicklung* (pp. 81–97). Weinheim: Juventa.

Roth, J. L., Malone, L. M., & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. *American Journal of Community Psychology*, *45*(3–4), 310–324. doi:10.1007/s10464-010-9303-3

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Sauerwein, M., Theis, D., & Fischer, N. (2016). How youths' profiles of extracurricular and leisure activity affect their social development and academic achievement. *International Journal for Research on Extended Education*, *4*(1), 103–124. doi:10.3224/ijree.v4i1.24778

Schüpbach, M. (2014). Extended education and social inequality in Switzerland: Compensatory effects? *Journal for Educational Research Online*, *6*(3), 95–114.

Schüpbach, M. (2015). Effects of extracurricular activities and their quality on primary school-age students' achievement in mathematics in Switzerland. *School Effectiveness and School Improvement*, *26*(2), 279–295. doi:10.1080/09243453.2014.929153

Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, *23*(4), 387–419. doi:10.1177/109821400202300403

Stecher, L., & Maschke, S. (2013). Research on extended education in Germany – A general model with all-day schooling and private tutoring as two examples. *International Journal for Research on Extended Education*, *1*(1), 31–52.

StEG-Konsortium. (2015). *Ganztagsschule 2014/2015: Deskriptive Befunde einer bundesweiten Befragung* [All-day schools 2014/2015: Descriptive findings of a nationwide survey]. Retrieved from http://www.projekt-steg.de/sites/default/files/StEG_Bundesbericht%202015_final_0.pdf

StEG-Konsortium. (2016). *Ganztagsschule: Bildungsqualität und Wirkungen außerunterrichtlicher Angebote: Ergebnisse der Studie zur Entwicklung von Ganztagsschulen 2012–2015* [All-day schooling: Education quality and effects of extracurricular activities: Results of the Study on the Development of All-Day Schools 2012–2015]. Retrieved from http://www.projekt-steg.de/sites/default/files/StEG_Brosch_FINAL.pdf

Steiner, C. (2011). Teilnahme am Ganztagsbetrieb: Zeitliche Entwicklung und mögliche Selektionseffekte [Participation in all-day schooling: Developments over time and possible selection effects]. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher, & I. Züchner (Eds.), *Ganztagsschule: Entwicklung, Qualität, Wirkungen: Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagsschulen (StEG)* (pp. 57–75). Weinheim: Juventa.

Steinhauer, H. W., & Zinn, S. (2016). *NEPS technical report for weighting: Weighting the sample of starting cohort 3 of the National Educational Panel Study (Waves 1 to 3)* (NEPS Working Paper No. 63). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXIII.pdf

Strietholt, R., Manitius, V., Berkemeyer, N., & Bos, W. (2015). Bildung und Bildungsungleichheit an Halb- und Ganztagsschulen [Educational outcomes and educational inequalities at half- and all-day schools]. *Zeitschrift für Erziehungswissenschaft*, *18*(4), 737–761. doi:10.1007/s11618-015-0634-6

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: A review journal of the Institute of Mathematical Statistics*, *25*(1), 1–21. doi:10.1214/09-STS313

Vandell, D. L., Reisner, E. R., & Pierce, K. M. (2007). *Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising afterschool programs*. Washington, DC: Policy Studies Associates.

Appendix 1. Descriptive covariate statistics for Sample 5–7 and Sample 7–9

	Sample 5–7 *N* = 1,131 students at 43 schools			Sample 7–9 *N* = 1,545 students at 64 schools		
	Source	Missing	*M*(*SD*) or % in categories	Source	Missing	*M*(*SD*) or % in categories
Mathematics pretest	student test Grade 5	5.4%	0.17 (1.11)	student test Grade 7	3.6%	0.93 (1.25)
Reading pretest	student test Grade 5	5.3%	0.15 (1.20)	student test Grade 7	3.5%	0.89 (1.38)
Latest school grade mathematics[1]	student quest. Grade 5	9.5%	3.86 (0.85)	student quest. Grade 7	11.4%	3.26 (0.97)
Latest school grade German[1]	student quest. Grade 5	9.9%	3.82 (0.81)	student quest. Grade 7	10.5%	3.35 (0.80)
School satisfaction[2]	student quest. Grade 5	7.2%	7.80 (2.41)	student quest. Grade 7	4.0%	6.95 (2.27)
Books at home[3]	student quest. Grade 5	10.1%	3.05 (1.32)	student quest. Grade 7	5.9%	3.11 (1.44)
Parent employment[4]	parent inter. Grade 5	32.2%	72.7% yes	parent inter. Grade 7	41.8%	78.7% yes
Homework help parents[5]	student quest. Grade 5	8.7%	2.68 (1.00)	student quest. Grade 7	8.1%	2.33 (1.02)
Homework help tutor[5]	student quest. Grade 5	14.6%	0.23 (0.76)	student quest. Grade 7	8.9%	0.52 (1.15)

Notes: Based on weighted imputed data; [1]from 0 = *failing* to 5 = *very good*; [2]from 0 = *completely dissatisfied* to 10 = *completely satisfied*; [3]from 0 = *0–10 books* to 5 = > *500 books*; [4]yes = "parent employed as a side-job, in part-time, or in full-time"; [5]from 0 = *never* to 4 = *always*.

Appendix 2. Treatment and control groups' comparison statistics of covariates in Sample 7–9

	Homework support			Remedial education			Subject-specific programs		
	M(*SD*) or % in categ. treated	*M*(*SD*) or % in categ. controls	*p* value group difference	*M*(*SD*) or % in categ. treated	*M*(*SD*) or % in categ. controls	*p* value group difference	*M*(*SD*) or % in categ. treated	*M*(*SD*) or % in categ. controls	*p* value group difference
Mathematics pretest	0.43 (1.19)	0.97 (1.25)	0.00	0.39 (1.15)	0.97 (1.25)	0.00	0.82 (1.33)	0.93 (1.25)	0.40
Reading pretest	0.33 (1.45)	0.95 (1.36)	0.00	0.48 (1.39)	0.93 (1.37)	0.01	0.68 (1.38)	0.91 (1.38)	0.22
Latest school grade mathematics	3.19 (1.03)	3.27 (0.96)	0.46	3.22 (0.93)	3.26 (0.97)	0.68	3.11 (1.08)	3.27 (0.96)	0.16
Latest school grade German	3.18 (0.88)	3.37 (0.79)	0.04	3.17 (0.82)	3.37 (0.79)	0.02	3.06 (0.83)	3.37 (0.79)	0.00
School satisfaction	6.23 (2.59)	7.01 (2.23)	0.01	6.48 (2.36)	6.99 (2.26)	0.02	6.32 (2.52)	6.99 (2.24)	0.01
Books at home	2.62 (1.42)	3.15 (1.43)	0.00	2.76 (1.50)	3.14 (1.43)	0.06	3.15 (1.49)	3.10 (1.43)	0.80
Parent employment	77.8% yes	78.8% yes	0.83	80.7% yes	78.5% yes	0.78	71.9% yes	79.1% yes	0.31
Homework help parents	2.27 (1.06)	2.33 (1.01)	0.62	2.35 (0.96)	2.33 (1.02)	0.85	2.32 (1.08)	2.33 (1.01)	0.91
Homework help tutor	1.25 (1.55)	0.45 (1.08)	0.00	0.97 (1.52)	0.48 (1.10)	0.00	0.83 (1.39)	0.50 (1.13)	0.09

Note: Based on weighted imputed data. N = 1,545 students at 64 schools.

Appendix 3. Variables in the imputation model

	Source	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
Student achievement variables						
Mathematics[1]	student test	X		X		X
Reading[1]	student test	X		X		X
Orthography[1]	student test	X		X		X
Perceptual speed[2]	student test	X				X
Cognitive reasoning[2]	student test	X				X
Declarative metacognition[3]	student test		X			X
ICT literacy[1]	student test		X			X
Scientific competence[1]	student test		X			X
Reading speed[2]	student test	X				X
All-day related variables						
Student use of						
Homework support[4]	student quest.		X		X	
Remedial education[4]	student quest.		X		X	
Subject-specific programs[4]	student quest.		X		X	
Subject-unrelated projects[4]	student quest.		X		X	
Leisure facilities[4]	student quest.		X		X	
Participation frequency[5]	student quest.		X		X	
School offer of						
Homework supervision[6]	principal quest.		X		X	
Remedial instruction[6]	principal quest.		X		X	
Subject-specific offering in mathematics[6]	principal quest.		X		X	
Subject-specific offering in German[6]	principal quest.		X		X	
Sports offerings[6]	principal quest.		X		X	
Music offerings[6]	principal quest.		X		X	
Free-time activities[6]	principal quest.		X		X	
Student background variables						
Latest school grade mathematics[7]	student quest.	X		X		X
Latest school grade German[7]	student quest.	X		X		X
School satisfaction[8]	student quest.	X		X		X
Books at home[9]	student quest.	X		X		X
Parent ISCED[10]	parent inter.	X		X		X
Partner ISCED[10]	parent inter.	X		X		X
Parent employment[11]	parent inter.	X		X		X
Partner employment[11]	parent inter.	X		X		X
Family native-born[12]	student quest.	X		X		
Language at home[13]	parent inter.	X		X		
Homework help parents[14]	student quest.	X		X		
Homework help tutor[14]	student quest.	X		X		
Gender[15]	cohort profile	X		X		
Age[16]	cohort profile	X		X		

Note: [1]Weighted maximum likelihood estimates; [2]sum scores; [3]mean scores; [4]yes = "participation"; [5]general participation in hours per week; [6]from 0 = *less than weekly offer* to 3 = *4–5 times per week*; [7]from 0 = *failing* to 5 = *very good*; [8]from 0 = *completely dissatisfied* to 10 = *completely satisfied*; [9]from 0 = *0–10 books* to 5 = > *500 books*; [10]from 0 = *inadequately completed general education* to 10 = *doctoral degree and postdoctoral lecture qualification*; [11]yes = "parent employed as a side-job, in part-time, or in full-time"; [12]yes = "child, parents, and grandparents born in Germany"; [13]yes = "language at home is other than German"; [14]from 0 = *never* to 4 = *always*; [15]0 = "male" and 1 = "female"; [16]in years.