Does Stimulating Physical Activity Affect School Performance?

Bart H. H. Golsteyn
Maria W. J. Jansen
Dave H. H. Van Kann
*Annelore M. C. Verhagen**

Abstract

This paper investigates whether encouraging children to become more physically active in their everyday life affects their primary school performance. We use data from a field quasi-experiment called the Active Living Program, which aimed to increase active modes of transportation to school and active play among 8- to 12-year-olds living in low socioeconomic status (SES) areas in the Netherlands. Difference-in-differences estimations reveal that while the interventions increase time spent on physical activity during school hours, they negatively affect school performance, especially among the worst-performing students. Further analyses reveal that increased restlessness during instruction time is a potential mechanism for this negative effect. Our results suggest that the commonly found positive effects of exercising or participating in sports on educational outcomes may not be generalizable to physical activity in everyday life. Policymakers and educators who seek to increase physical activity in everyday life need to weigh the health and well-being benefits against the probability of increasing inequality in school performance.

INTRODUCTION

Due to the global increase of childhood obesity and sedentary behavior, governments and institutions advocate increasing the amount of time primary school children spend on physical activity.[1] Although the benefits of physical activity for physical and mental health are well-documented, the current literature is still unsettled regarding the causal effects on educational outcomes. Physical activity is expected to positively affect cognition and thereby school performance from a physiological viewpoint,[2] but the effect of increasing physical activity on school performance is

*Author names are listed in alphabetical order.
[1] See, e.g., Centers for Disease Control and Prevention (2011); European Commission (2014); World Health Organization (2010).
[2] The medical literature shows that physical activity can positively affect cognitive functioning through increased blood and oxygen flow to the brain and increased growth factors that help create new nerve

Journal of Policy Analysis and Management, Vol. 39, No. 1, 64–95 (2020)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.22156

theoretically ambiguous. Even if physical activity has positive effects on children's cognitive abilities, increasing sports participation may decrease focus and attention during instruction periods, or crowd out time investments in other potentially beneficial activities such as studying or active play.[3] Physical activity may, however, also crowd out potentially harmful activities such as smoking or watching television, leading to a zero or positive effect on school performance. The net effect on school performance thus depends on the relative gain (or harm) from the activity that was crowded out compared to the gains from physical activity.

This paper investigates with a field quasi-experiment whether the encouragement of informal physical activity, such as active play, affects primary school performance. The quasi-experiment, called the Active Living Program, is organized in low-SES areas of the Netherlands. We analyze whether the interventions affect physical activity and school performance.

The Active Living Program aims to increase physical activity and decrease sedentary behavior among 8- to 12-year-olds, through encouraging active modes of transportation to school and active play at school and during leisure time. The interventions, which were implemented between April 2013 and June 2014, did not affect instruction time or the structure or content of classes. Pre- and post-treatment data regarding school performance and time spent on physical activity were collected in 10 treatment schools and 11 control schools. School performance is measured through nationally standardized language and math/calculating tests that use a grading scheme common to all schools. Physical activity data were collected both pre- and post-treatment by means of accelerometers that children wore for five consecutive days. We use a difference-in-differences technique to estimate the effects of the Active Living Program on school performance and time spent on physical activity.

The results indicate that the Active Living Program causes a significant decrease in school performance. Our most conservative estimate of the average treatment effect is −5.9 percent of a standard deviation. The negative effect on school performance is strongest among the worst-performing students. Event study analyses including three pre-treatment years and several robustness checks provide strong indications that the effects on school performance are indeed due to the interventions in the treatment schools, which allows for a causal interpretation of our estimates. We explore which mechanisms may account for our main findings. First, we find that the Active Living Program causes a significant increase in time spent on physical activity during school hours (0.34 standard deviations or 9.30 minutes per day), but no significant effect on leisure-time activity. Second, we find that the Active Living Program reduces children's self-assessed ability to be calm and quiet when the teacher wants them to (i.e., increased restlessness). The self-assessed ability to be calm and quiet is positively related to school performance. Although the results must be interpreted with caution, they imply that an additional half-hour spent on physical activity during school hours, decreases school performance by 0.19 standard deviations. Approximately half of this effect appears to be driven by increased restlessness during instruction time.

cells and support synaptic plasticity, decreased stress, and improved mood due to higher levels of norepinephrine and endorphins (Singh et al., 2012). Erickson, Hillman, and Kramer (2015), for example, find that more-active children show greater hippocampal and basal ganglia volume, greater white matter integrity, elevated and more efficient patterns of brain activity, superior cognitive performance, and higher scholastic achievement.

[3] As Lizandra et al. (2016) point out, sedentary activities could either be academic (e.g., doing homework, reading, or studying), technology-based (e.g., playing video games or watching television) or social-based (e.g., sitting with friends or chatting via social networks); each of these behaviors has a different expected effect on academic performance.

This paper contributes to the literature that studies causal effects of physical activity interventions on educational outcomes of children and adolescents. Earlier papers focus on the effect of participating in physical education classes, club sports, or other forms of formal or professional sports (e.g., Barron, Ewing, & Waddell, 2000; Cabane, Hille, & Lechner, 2016; Dills, Morgan, & Rotthoff, 2011; Felfe, Lechner, & Steinmayr, 2016; Lipscomb, 2007; Pfeifer & Cornelissen, 2010; Rees & Sabia, 2010; Stevenson, 2010). Our study contributes to this literature by being the first to analyze the effect of the encouragement of informal activities such as active transportation and active play. Moreover, due to the quasi-experimental design, our identifying assumptions are less restrictive than in prior studies.

Our results suggest that the commonly found positive effects of exercising or participating in sports on educational performance may not be generalizable to physical activity in everyday life. This has important implications for policymakers, educators, and researchers in other high- or middle-income countries, who aim to increase physical activity among children and adolescents within or outside of low-SES regions. Our findings highlight that: (1) policymakers must remain alert when implementing interventions and assure that all the (assumed) benefits and potential costs are carefully monitored; (2) policymakers must consult research evidence to make informed decisions that weigh the health and well-being benefits against the potential increased inequality in school performance; and (3) schools may need targeted interventions that increase low-performing students' physical activity without further decreasing their educational performance.

This paper proceeds as follows. The next section describes our contribution to the literature. The third section presents the empirical strategy and discusses the field quasi-experiment. The fourth section describes the data. The fifth section provides the main results, and the sixth section provides robustness checks. The seventh section analyzes the effects of the Active Living Program on time spent on physical activity and investigates the external validity of our findings. The eighth section analyzes the mechanisms. The ninth section concludes.

CONTRIBUTION TO THE LITERATURE

Earlier papers on the consequences of physical exercise mostly have shown positive effects on health (see, e.g., Cawley, Frisvold, & Meyerhoefer, 2013; Cawley, Meyerhoefer, & Newhouse, 2007; Strong et al., 2005) and on cognitive functioning, particularly in older adults (see, e.g., Colcolmbe & Kramer, 2003; Hillman, Erickson, & Kramer, 2008). A limited set of papers analyzes the causal effect of physical activity (PA) on educational outcomes, such as school performance, for younger individuals.

When analyzing the effect of PA on school performance, there is a risk of reverse causality or selection based on unobserved factors. For example, unobserved health deterioration may cause a simultaneous decrease in school performance and time spent on PA. Another example is that children with decreasing school performance may be encouraged to spend more time doing homework: a sedentary activity. Previous attempts to overcome such problems include adopting an individual-fixed-effects approach (Lipscomb, 2007; Rees & Sabia, 2010), and selection-on-observables or matching (Cabane, Hille, & Lechner, 2016; Felfe, Lechner, & Steinmayr, 2016; Pfeifer & Cornelissen, 2010). Other studies adopt an instrumental variable approach in which students' sports participation is instrumented by their height (Barron, Ewing, & Waddell, 2000; Pfeifer & Cornelissen, 2010; Rees & Sabia, 2010), by certain parental characteristics such as income, or by school characteristics such as school size and the books-per-student ratio (Barron, Ewin, & Waddell, 2000). Stevenson (2010) and Dills, Morgan, and Rotthoff (2011) use changes in state-level policies to instrument for changes in sports participation at school.[4]

Our focus on informal PA makes several important contributions to the literature. First, being physically active entails more than just dedicated physical exercise. During a day, the amount of time spent on informal PA is (or can be) substantially larger than the amount of time spent on dedicated physical exercise. For several reasons, it is unclear whether we should expect the effects of the encouragement of informal PA to be similar to the effects usually found for formal PA. First, given that informal PA is usually less intense than participation in sports, our study provides more insight into the dose-response relationship between PA and school performance. Previous research has shown that the positive relation between PA on health and educational outcomes may be driven mostly by moderate-intensity and vigorous PA (see, e.g., Felez-Nobrega et al., 2017; Singh et al., 2012; Van den Berg et al., 2016). Second, formal and informal PA may foster the development of different skills. Sports teach players competitiveness and perseverance—skills that could be beneficial for educational outcomes, but which are likely less pronounced (if present at all) in active play. Additionally, athletes may receive increased attention and encouragement from their teachers or parents, leading to increased self-esteem and increased peer pressure to succeed (Stevenson, 2010). Developing these "sports-related skills" implies that potential crowding-out effects will be less pronounced. In other words, if children reallocate study hours to PA, the effect of engaging in sports on school performance might have been negative had it not been for the athlete's development of these additional sports-related skills.

Our research also contributes to the literature in that in our quasi-experiment, unlike other PA intervention studies, none of the interventions affected school time or the curriculum. Moreover, our PA data are collected by accelerometers that children wore throughout the day, which allows us to capture any substitution or complementarity effects between in- and out-of-school PA, and between formal and informal forms of PA. This is an important benefit relative to other studies, which focus on (self-reported) measures of formal PA.

EMPIRICAL STRATEGY

The Active Living Program[5]

In order to stimulate PA, we organized the Active Living Program: a field quasi-experiment that took place between 2012 and 2014 in 21 primary schools located in low-SES areas in the Southern-Limburg region of the Netherlands[6] (see Figure D1 in Appendix D for a map of the research area).[7] The target group consists of children who were enrolled in grades 4 or 5 during the 2012/2013 school

[4] Stevenson (2010) finds a positive effect of increased high school athletic participation on education and labor market outcomes for women. Dills, Morgan, and Rotthoff (2011) do not find any significant effects of increased recess or physical education time on school performance.

[5] The text in this section is partly taken from van Kann et al. (2016). For a more detailed description of the study design of the Active Living Program and its methods and implementation strategies, see van Kann et al. (2015).

[6] The Limburg region has about 609,000 inhabitants and a population density of 922 inhabitants per square kilometer. This density is almost twice as high as the average population density in the Netherlands, which is 496 inhabitants per square kilometer. The region consists of 18 municipalities. The average disposable household income in the region is about €31,500 per year, which is somewhat lower than the national average of €34,200 per year (Statistics Netherlands, 2013).

[7] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

year and transferred to grades 5 or 6 during the 2013/2014 school year (i.e., 8- to 12-year-olds).

We focus on 8- to 12-year-old children because children in this age category experience more independent mobility due to fewer parental restrictions than younger children. For instance, they can independently use transportation or visit a playground in their neighborhood. Independent mobility has been shown to be an important predictor of children's PA (Carver et al., 2013; Hume et al., 2009; Pont et al., 2009). Moreover, independent mobility implies that interventions can be directly targeted at children rather than indirectly influencing their behavior through their parents.

Ten primary schools agreed to participate as treatment schools.[8] We matched each treatment school to a control school based on the level of neighborhood deprivation and the level of urbanization (urban vs. rural). One treatment school was matched to an additional control school because the school to which it was initially matched relocated during the 2012/2013 school year. After the pre-treatment PA measurement, no further changes were observed in this control school, making both control schools eligible for study inclusion. We exclude one treatment school from the data because it organized a sports day during the pre-treatment PA measurement, resulting in exceptionally high levels of PA that were not representative of children's average school days.

Each treatment school had a working group that consisted of representatives of the school, parents, municipal authorities and other stakeholders, and a public health services (PHS) employee who served as chair. The researchers trained the PHS employee in "physical and social environmental thinking" and evidence-based PA intervention opportunities. The working groups were responsible for choosing, designing, and implementing intervention elements, but the research team had final approval over the final plan costs. Although the overall scope was the same, the intervention packages differed in magnitude and design across treatment schools, depending on local needs.

All working groups received an intervention budget of €2,000 at the start of the project. However, costs for environmental adaptations in public spaces (i.e., school surroundings) were covered by municipalities, and if intervention plans exceeded the budget, the working group could apply for additional funding from the research team and municipal authorities (which had a representative as a member of the working group). Additionally, the province of Limburg provided funding for a sports day at all treatment schools, and several treatment schools organized charity events such as a charity run in order to raise money for new playground equipment. Treatment schools could implement the interventions as soon as the pre-treatment PA measurement was conducted at their school (see the section on Data below for more information on the PA data collection). All interventions were implemented before the second PA measurement in the spring of 2014. Table 1 presents an overview of all interventions per treatment school. Table D1 in Appendix D provides a more detailed description of the implemented interventions per treatment school.[9]

Each intervention was intended to decrease sedentary behavior or increase PA at school or during leisure time by nudging children to become more active. These

[8] See Textbox D1 and Figure D2 in Appendix D for a more detailed description of how treatment and control schools were selected. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

[9] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 1. Overview of implemented interventions per treatment school.

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Panel A. Active transportation to school									
Develop safe route to school	−	−	+	+	+	−	−	−	+
Mobilize crossing guards	−	−	+	+	−	−	−	−	−
Adapt unsafe intersection in school environment	−	−	+	−	−	−	−	−	−
Make bicycle racks available	−	+	−	−	−	+	−	−	−
School pedestrian crossing indicators	+	+	−	−	−	+	−	−	−
Create safer parking situation around school	−	−	+	−	−	−	−	−	−
Create traffic circle in schoolyard environment	+	+	−	−	−	−	−	−	−
Sticker competition for active transport	+	+	+	+	−	−	+	−	−
School stimulation documentation on safe active transport	−	−	+	+	−	−	−	−	−
Introduce "Walk/Bike-to-school-day"	−	−	−	−	−	+	−	−	−
Speed-check action performed by children	−	−	+	+	−	−	−	−	−
Lessons to improve bicycle skills	−	−	−	−	+	−	−	−	+
Panel B. PA in school									
New fixed equipment in schoolyard	−	+	−	+	+	+	−	+	+
New loose equipment in schoolyard	−	+	+	−	+	+	+	−	+
Playground markings	−	+	−	−	−	+	+	−	−
Establish ball game area	+	+	−	−	−	+	−	+	−
Put a ball backstop beside railway	+	−	−	−	−	−	−	−	−
Sound equipment in schoolyard environment	−	+	−	−	−	−	−	−	−
Additional sports day in schoolyard	+	+	+	+	−	−	−	+	−
Sports clinics at recess	+	+	+	+	+	+	−	−	+
Use of schoolyard games	−	+	+	−	−	−	+	+	−
Contest for best PA encouragement idea	+	−	−	−	−	−	−	+	−
Panel C. PA in leisure time									
Establish training circuit	−	−	+	−	−	−	−	−	−
Active Living Games	+	+	+	+	+	+	+	−	+
Establish out-of-school PA program	+	+	+	+	+	−	−	+	+
Establish school soccer team	−	−	+	−	−	−	−	−	−
Establish PA activities by children for local residents	−	−	+	−	−	−	−	−	−

Notes: Tn = Treatment school (number); + = intervention was implemented; − = intervention was not implemented. One of the initial 10 treatment schools is excluded from the analyses because it organized a sports day during the pre-treatment physical activity measurement week. Control schools did not implement any interventions. See Table D1 in the Appendix for a detailed explanation of each intervention.[10]

nudges could be explicit, e.g., by organizing additional sports days, but also could be implicit, e.g., by creating safer traffic environments around the treatment schools.[11] It is important to note that none of the interventions affected school time or the structure or content of the curriculum because this may have otherwise affected school performance. For the same reason, interventions requiring structural changes in the school yard (e.g., installing new fixed equipment, a ball backstop, or sound equipment in the school yard) were implemented before or after school hours.

[10] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[11] The only implemented interventions that were specifically aimed at increasing time spent on group PA were the "Active Living Games" and the establishment of a school soccer team. All other interventions are not specifically aimed at either individual or group PA. For example, children could choose whether they wanted to play with the new playground equipment with others or by themselves.

The intervention budget of €2,000 euros was not conditional on actual implementation of interventions. However, the research team visited the treatment schools regularly pre- and post-treatment, which allowed us to confirm that changes had been made to the physical school environment. Picture D1 in Appendix D provides examples of the interventions.[12] Children were not obliged to engage in additional PA. All implemented interventions were intended to nudge children toward increased PA time (i.e., intention to treat), but there was no guarantee that children increased their time spent on PA.

To decrease the probability of contamination of the control group, all interventions were implemented within an 800-meter radius around the treatment schools. In the Netherlands, the majority of the primary school children live within 800-meters of the school (Statistics Netherlands, 2014). Control schools are located outside of the 800-meter radius around treatment schools, and they did not implement any Active Living interventions. When conducting the pre- and post-treatment PA measurements, the research team confirmed that control schools did not make any changes to the physical school environment. However, we cannot completely rule out the possibility that control schools changed something between the pre- and post-treatment year that had an effect on the children's time spent on PA. We therefore conduct robustness checks in which we analyze whether there was an unexpected treatment effect in the control schools (see the section on Robustness Checks below).

Identifying the Treatment Effect on School Performance

After the data collection (see the next section for more information), we verified whether treatment and control schools were similar with respect to pre-treatment observables. These analyses revealed that before any interventions were implemented, treatment and control schools did differ in various domains: the most important differences were that children in treatment schools performed significantly better on math tests and spent an average of 15 minutes per day less on PA during school hours (see Table D2 in Appendix D).[13] This means that matching appears not to have been successful. Therefore, we could not merely compare post-treatment outcomes between children in treatment and control schools because post-treatment differences may be due to self-selection.

When there are pre-treatment differences between the treatment and control groups, a difference-in-differences technique can be used as an alternative strategy for identifying treatment effects. Underlying the technique is the identifying assumption that the initial differences between the treatment and control groups would have remained similar if the Active Living Program had not been implemented. Common trends in the pre-treatment period suggest that this assumption holds (see the next subsection).

In order to identify the effect of the Active Living Program on school performance, we estimate the following model:

$$Score_{ist} = \alpha_0 + \alpha_1 \left(Treat_{is} * Post_t\right) + \alpha_2 Post_t + \alpha_3 age_{ist} + \alpha_4 age^2_{ist} + \gamma_{1is} + \varepsilon_{1ist}. \quad (1)$$

In equation (1), i indexes the individual pupil, s indexes the school, and t indexes time. In the regressions, *Treat* is an indicator variable with a value of one if a pupil

[12] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[13] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

is in a school that belongs to the treatment group and zero if he or she is enrolled in one of the control schools.[14] *Post* is an indicator variable with a value of one for tests taken during the post-treatment year of 2013/2014 and zero for tests taken during the pre-treatment year of 2012/2013.[15] We are interested in the difference-in-differences estimator α_1, i.e., the interaction of *Treat* and *Post*. In our main specification, we control for age and child fixed effects γ.[16]

Common Trends: Event Study Analysis

To analyze whether pre-treatment trends are indeed parallel, we conduct an event study analysis including three pre-treatment years by estimating the following model:

$$Score_{ist} = \alpha_5 + \alpha_6\,(Treat_{is} * Post_t) + \alpha_7\,(Treat_{is} * Pre2_t) + \alpha_8\,(Treat_{is} * Pre3_t) + \delta_t + \alpha_9 age_{ist} + \alpha_{10} age_{ist}^2 + \gamma_{2is} + \varepsilon_{2ist}. \quad (2)$$

The variables *Post*, *Pre2*, and *Pre3* are indicator variables for tests taken during the school years 2013/2014, 2011/2012, or 2010/2011, respectively (we use the 2012/2013 school year, i.e., *Pre1*, as the reference group). δ_t stands for year fixed effects, and similar to equation (1), we control for age and child fixed effects. The treatment effect in the treatment year is given by the interaction between *Treat* and *Post* (α_6), which should be almost identical to α_1 in equation (1). The parallel trends assumption is supported if α_7 and α_8, i.e., the estimated treatment effects in the pre-treatment years, are both not significantly different from zero. We formally test whether these two estimators are individually and jointly significantly different from zero.

Robust Standard Errors

We use robust clustered standard errors because school performance may be correlated among pupils within the same school. However, only 21 schools participated in the Active Living Program, which creates a risk that our robust standard errors will be substantially downward biased (see, e.g., Angrist & Pischke, 2008; Bell & MacCaffrey, 2002; Bertrand, Duflo, & Mullainathan, 2004; Cameron, Gelbach, & Miller, 2008; Cameron & Miller, 2015; Donald & Lang, 2007).

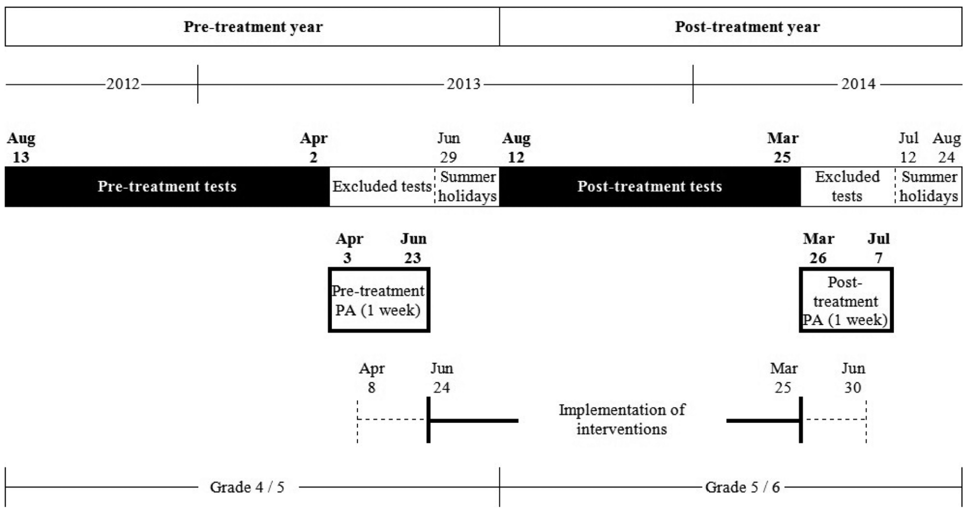
To overcome this problem, we use the method proposed by Donald and Lang (2007), which has been shown to work reasonably well in difference-in-differences settings. They suggest that inferences be based on a *t*-distribution with degrees of freedom equal to the number of clusters minus the number of variables that are constant within clusters, rather than the standard normal distribution. By doing so, one essentially recognizes that the fundamental unit of observation is a cluster and not an individual within a cluster. We verified that inference based on this approach is indeed the most conservative compared to no clustering, clustering without adjusted critical *t*-values, or clustering with corrected standard errors based

[14] The sample did not include any children who changed schools between the pre- and post-treatment year.
[15] Note that the interventions could be implemented throughout the 2013/2014 school year (March through June 2014), which implies that some tests taken during this school year may be wrongly defined as post-treatment tests. However, as this would attenuate results, we consider it a minor issue.
[16] It is possible to include both age and year fixed effects because in our sample, age is measured in months when the tests are taken, and test dates vary by year. However, note that including child fixed effects implies that we cannot also include *Treat* because this variable does not vary across time.

on the Moulton factor (as proposed by Angrist & Pischke, 2008).[17] Furthermore, in order to avoid serial correlation, we compare the pre-treatment year (2012/2013) to the post-treatment year instead of using test scores in all pre-treatment years as the baseline measure (Bertrand, Duflo, & Mullainathan, 2004).

DATA

Our analyses are based on two data sets. The first contains data on individual test score histories from children in nearly all schools in the Southern Limburg region (school performance data). The second contains data on time spent on PA as measured by accelerometers (PA data) worn by children in the 21 schools that participated in the Active Living Program (Active Living schools). We discuss these data sets in turn. See Figure 1 for a timeline of the data collection process.



Notes: In each school, PA data were collected during one week between April and June 2013 and again during one week between March and July 2014. Treatment schools could implement their interventions between the last day of the pre-treatment PA measurement week in 2013 and the first day of the post-treatment PA measurement week in 2014. Implementation periods can therefore differ by school as indicated by the dashed lines. In some analyses, pre-treatment test scores taken before August 13, 2012 are also included.

Figure 1. Timeline of the Data Collection.

School Performance Data

The data on school performance are collected from over 200 schools in a cooperative project between elementary schools, school boards, municipalities, and Maastricht University, called the Onderwijs Monitor Limburg. The 21 Active Living schools are included in these 200 schools. We use data on the school performance of children in the schools that did not participate in the Active Living Program as a treatment or control group (non-Active Living schools) in the robustness checks.

[17] In some cases, inference based on one of the other strategies yields slightly more conservative results, but in these cases, the interpretation of the results does not change, i.e., the rejection of the null hypothesis stands.

The school performance data include a wide variety of standardized tests that children took between Kindergarten and grade 6 (i.e., ages 4 to 12) in various domains such as reading, spelling, vocabulary, calculating, and math. The tests are graded by the teachers who use a grading scheme that is standardized across all schools in the Netherlands. Each subject has its own range of scores. For each subject, there are two tests per grade with similar content. The teacher can decide whether a child takes both tests and whether the child takes the same test more than once.

For the purpose of this study, we limit our sample to the target group of the Active Living Program, i.e., children who were enrolled in grades 4 or 5 in 2012/2013 and in grades 5 or 6 in 2013/2014.[18] We exclude tests that were taken during the PA measurements to prevent reactivity effects.[19] Consequently, we exclude all tests that were taken between April 3 and August 11, 2013, and between March 26 and August 24, 2014.[20]

Because children can take multiple tests of the same subject within a school year, we first aggregate children's subject-scores to the average within each school year.[21] To create a comparable scale across subjects (as each has their own range of scores), we standardize children's average subject-scores across schools (Active Living schools as well as non-Active Living schools) and school years to a mean of zero and a standard deviation of one.

We can then construct an overall school performance variable by calculating children's average score across their standardized subject-scores within a school year. For interpretation purposes, we rescale this variable on a scale from zero to 100. This is the dependent variable in our main analyses on school performance. We repeat this strategy for language and math tests separately.

PA Data

The school performance data can be merged at the individual level with the PA data. Time spent on PA was measured during one week between April 3 and June 24, 2013 and again during one week 12 months later (March 26 to July 7, 2014). Prior to the pre-treatment PA measurement, 61.6 percent of the children in the target group obtained written parental consent to wear an accelerometer. We verified that these children were not significantly different with respect to observable characteristics compared to those for whom we did not receive parental consent. In the pre-treatment year, 791 children were fitted with an accelerometer, which they were asked to wear on their waist for at least five consecutive days, including at least one weekend day.[22] To correct for potential seasonal effects (see, e.g., Rich, Griffiths, & Dezateux, 2012), we assessed every treatment-control school pair on the same dates.

PA data were collected using accelerometers (ActiGraph GT3X+; 30 Hz, 10-second epochs). The accelerometers' sensors measure acceleration in units of gravity on

[18] In our sample, no children repeated grade 4 in 2013/2014. We include in the analyses children who repeated grade 5 in 2013/2014, as well as those who skipped grade 5 and went directly from grade 4 in 2012/2013 to grade 6 in 2013/2014.

[19] We also exclude the few tests that were taken between the last PA measurement and the summer holidays. No tests are taken during the summer holidays.

[20] The results from analyses including these tests remain qualitatively robust.

[21] Around 80 percent of the tests are taken in January and February (school years run from August until June), both in the pre- and post-treatment years. Therefore, we cannot estimate the treatment effect non-parametrically in the post-treatment period, e.g., by month, due to a lack of observations in all other months.

[22] Parental consent was requested of parents of 1,322 children and obtained for 815 children (61.6 percent). Out of the 815 children for whom we obtained parental consent to wear an accelerometer, 791 were fitted with an accelerometer in the pre-treatment year (97 percent); 24 children did not wear their accelerometer.

three axes: vertical, horizontal, and perpendicular. The accelerometers run on a battery and are not turned on or off by the participants. They register acceleration continuously. We use Evenson's cutoffs (Evenson et al., 2008) to determine PA intensity levels: accelerations of more than 100 counts per minute are recorded as time spent on PA and accelerations of 100 counts per minute or less are recorded as time spent on sedentary behavior.[23] We distinguish between sedentary behavior and a child's not-wearing the accelerometer based on Choi's classification criteria (Choi et al., 2011). We use the ActiLife software (version 6.10.4, ActiGraph, Pensacola, U.S.A.) to transform the raw data into data per hour.

"School-time PA" is defined according to regular school times of Dutch primary schools, i.e., Mondays, Tuesdays, Thursdays, and Fridays from 9:00 a.m. until 3:00 p.m. and Wednesdays from 9:00 a.m. until noon. Leisure time is defined as the hours before and after school time and the weekends. PA data collected between 11:00 p.m. and 6:00 a.m. are excluded from the analyses, because these data are defined as sleeping hours.[24] The PA data are collected per hour, per day. In our main analyses, we sum the PA data per day per part (school time/leisure time) for each child.

We set a minimum amount of accelerometer wearing time, as otherwise, we would have to assume, for example, that if a child wore the accelerometer for only five minutes during an entire day, those five minutes are representative of the child's activity levels throughout that day. We define minimum wear time as at least half the available daily amount of school/leisure time.[25] This means that, regarding school-time PA, minimum wear time is defined as at least 180 minutes (three hours) between 9:00 a.m. and 3:00 p.m. on Mondays, Tuesdays, Thursdays, and Fridays, and at least 90 minutes (1.5 hours) between 9:00 a.m. and noon on Wednesdays.[26] Minimum wear time for leisure-time PA is defined as at least 330 minutes (5.5 hours) on Mondays, Tuesdays, Thursdays, and Fridays, at least 420 minutes (seven hours) on Wednesdays, and at least 480 minutes (eight hours) on Saturdays and Sundays.[27]

We aggregate the children's daily time spent on PA (school-time/leisure-time) to the average amount of time spent on PA per measurement week. This means that in our analyses, PA is measured in average number of minutes per day (per measurement week).

Summary Statistics

In our estimation samples, we have school performance data from 1,014 children, school-time PA data from 536 children, and leisure-time PA data from 509 children. Around half of the sample is enrolled in a treatment school. On average, children took five tests per year, with an average score of 57.28 in the pre-treatment year and 65.16 in the post-treatment year.

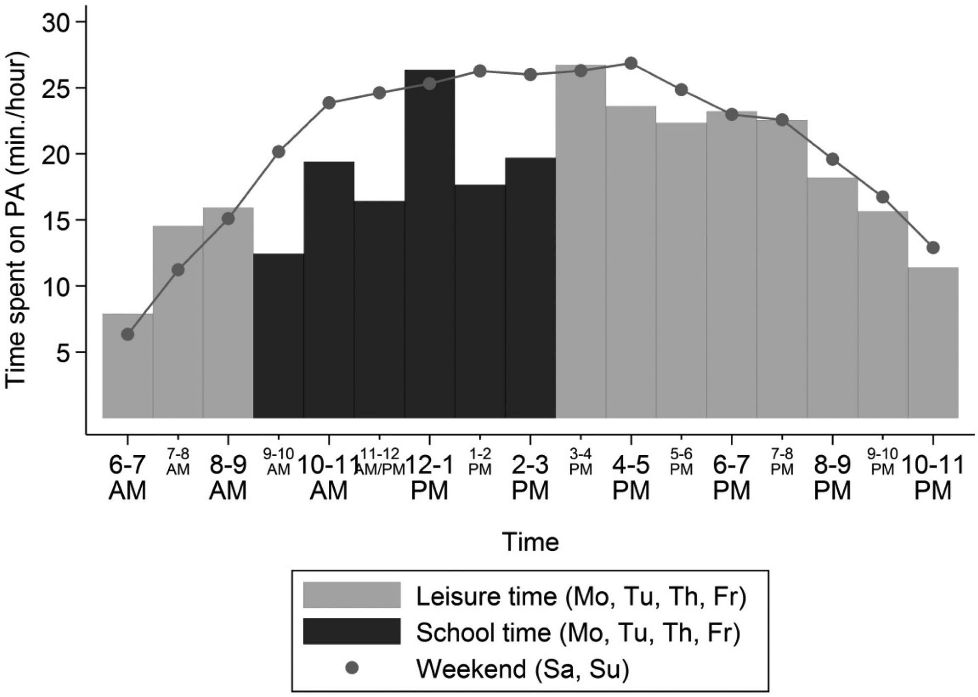
[23] Counts per minute is a unit of activity that is commonly used in health sciences. Light intensity PA (e.g., walking, biking slowly, playing catch) is defined as 101 to 2,295 counts per minute, 2,296 to 4,011 counts per minute is defined as moderate intensity PA (e.g., brisk walking, jumping on a trampoline, recreational swimming), and 4,012 or more counts per minute is defined as vigorous intensity PA (e.g., running, jumping rope, swimming laps). Fewer than 100 counts per minute is defined as sedentary behavior, e.g., watching TV, gaming, or doing homework (Evenson et al., 2008).

[24] Usually, these data are already excluded because they are recognized as non-wearing time (Choi et al., 2011). Excluding the 11:00 p.m. to 6:00 a.m. data reduces the number of child-hour observations by 1.91 percent.

[25] In our analyses per hour, we define minimum wear time as at least 30 minutes per hour.

[26] This reduces the number of child-day-school-time observations by 1.89 percent. The results are qualitatively robust to this exclusion.

[27] This reduces the number of child-day-leisure-time observations by 11.58 percent. The results are qualitatively robust to this exclusion.



Source: PA data.

Figure 2. Pre-Treatment Time Spent on PA (Excluding Wednesdays).

Children wore their accelerometer for four days on average during each PA measurement week. During these days, they wore their accelerometer for approximately five hours per day during school hours (309 minutes) in the pre-treatment year, out of which they spent 1.5 hours (93 minutes) on PA. Figure 2 shows that most of the school-time PA (black bars) takes place between noon and 1:00 p.m., which is also the time when most schools have recess. During leisure time in the pre-treatment year, children wore their accelerometer for approximately nine hours per day (548 minutes), out of which they spent 3.5 hours (204 minutes) on PA. Figure 2 shows that time spent on PA during leisure time does not differ substantially between school days (grey bars) and weekend days (dotted line). However, the figure does highlight the “incarceration effect” of school on PA; children are much more active between 9:00 a.m. and 3:00 p.m. on weekend days than during school days.

Between the pre- and post-treatment year, average time spent on PA during school time did not change significantly, but average leisure-time PA decreased by 21 minutes per day.[28] Appendix A describes the control variables we use in the analyses and provides summary statistics of all variables.[29]

[28] Note that this trend is controlled for in the difference-in-differences estimations.
[29] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher’s website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

RESULTS

We first show the “naïve” OLS estimates of the relationship between time spent on PA and school performance before the Active Living interventions were implemented, i.e., in 2012/2013. The results in Table 2 show that, in our sample, every additional 10 minutes spent on PA during school time is related to a 0.52 point lower test score (5.6 percent of a standard deviation in the full sample). This negative correlation between school-time PA and school performance is significantly different from zero at the 1 percent level. Pre-treatment leisure-time PA is not significantly related to pre-treatment school performance.

However, these analyses could be biased due to reverse causality or omitted variables. It is possible that children who spend less time on PA at school use that time to invest in other cognitively stimulating activities such as studying, reading, or playing music. To account for this, we will now turn to our difference-in-differences estimations based on the Active Living Program (see Table 3).

The difference-in-differences estimations reveal that the Active Living Program has a negative effect on school performance, which is significantly different from zero at the 5 percent level. Due to the interventions, the increase in school performance of children in the treatment schools is 1.17 points smaller (0.13 standard deviations) compared to the increase of those in the control schools. This result is robust across all specifications, i.e., without control variables, including child fixed effects, controlling for age as well as child fixed effects, and when we adjust p-values for the small number of clusters. Additional analyses reveal that the effect of the Active Living Program is similar for school performance on language tests (-0.12 standard deviations) and math tests (-0.15 standard deviations; see Table D3 in Appendix D).[30]

Next, we analyze whether the effect of the Active Living Program on school performance is different for the worst-performing students compared to the best-performing students. We estimate these heterogeneous effects nonlinearly, using a moving window of 200 individuals across the school performance distribution in the pre-treatment year of 2012/2013. For each selection of 200 individuals, the estimated treatment effect on school performance is depicted on the y-axis of Figure 3. The figure indicates that the Active Living Program negatively affects school performance across the distribution. However, the negative effect is strongest in the left tail of the distribution, i.e., among those who were performing worse at school during the pre-treatment year compared to their peers.

To analyze whether the treatment effects persist over time, we estimate the treatment effect one additional year post-treatment, i.e., in 2014/2015. An important point is that we can only estimate this long-term treatment effect for half our sample, namely for those who were enrolled in grade 4 in 2012/2013 and were (therefore) enrolled in grade 6 in 2014/2015. Those who were enrolled in grade 5 in the pre-treatment year of 2012/2013 finished primary school in 2013/2014 and were enrolled in secondary education in 2014/2015. The results from these analyses reveal that the point estimate remains similar in the longer run (2014/2015). However, the estimates in both 2013/2014 and in 2014/2015 are not statistically significant, which may be due to the smaller sample size (see Figure D3 in Appendix D).[31]

[30] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher’s website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[31] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher’s website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 2. Pre-treatment correlation between time spent on physical activity and school performance.

	(1)	(2)	(3)	(4)	(5)	(6)
	PA during school time			PA during leisure time		
Dependent variable: overall school performance	Base	Incl. controls from school perf. data	Incl. controls from PA data	Base	Incl. controls from school perf. data	Incl. controls from PA data
Time spent on PA (min./day)	−0.033	−0.031	−0.052	−0.008	−0.007	−0.009
	(0.013)	(0.011)	(0.012)	(0.005)	(0.005)	(0.005)
Adjusted p-value	[0.018]	[0.012]	[0.004]	[0.142]	[0.214]	[0.116]
Age when tests were taken (months)		0.966	0.969		1.112	0.938
		(0.783)	(0.704)		(0.778)	(0.595)
Age squared		−0.005	−0.005		−0.005	−0.004
		(0.003)	(0.003)		(0.003)	(0.002)
Cohort (1 = 5th graders, 0 = 4th graders)		10.183	10.096		10.080	9.657
		(0.819)	(0.879)		(0.957)	(1.032)
Gender (1 = boy)		0.089	0.326		0.193	0.255
		(0.981)	(0.962)		(1.100)	(1.027)
Wearing time of accelerometer (min./day)			0.013			0.129
			(0.101)			(0.050)
Wearing time squared			−0.000			−0.000
			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			−0.054			−0.052
			(0.031)			(0.040)
Temperature squared			0.000			0.000
			(0.000)			(0.000)

Table 2. Continued.

Dependent variable: overall school performance	(1)	(2) PA during school time	(3)	(4)	(5) PA during leisure time	(6)
	Base	Incl. controls from school perf. data	Incl. controls from PA data	Base	Incl. controls from school perf. data	Incl. controls from PA data
Amount of sunshine during PA measurement week (hours/day)			1.125			0.591
			(0.530)			(0.659)
Sunshine squared			−0.158			−0.081
			(0.085)			(0.101)
Rain during PA measurement week (1 = yes)			−0.562			−0.507
			(0.924)			(1.011)
Constant	60.654	9.015	9.258	59.427	−2.588	−26.163
	(1.393)	(47.984)	(49.228)	(1.046)	(48.199)	(43.406)
Observations	549	549	512	533	533	500
R-squared	0.012	0.270	0.288	0.003	0.263	0.281
Mean school performance in estimation sample	57.665	57.665	57.668	57.743	57.743	57.767
Standard deviation	(7.795)	(7.795)	(7.687)	(7.827)	(7.827)	(7.681)
Mean school performance in full sample	57.803	57.803	57.803	57.803	57.803	57.803
Standard deviation	(8.304)	(8.304)	(8.304)	(8.304)	(8.304)	(8.304)

Notes: The table shows the results of six ordinary least squares regression analyses. The dependent variable in each column is overall school performance in 2012/2013. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: Merged school performance and PA data.

Table 3. Difference-in-differences estimations of the effect of the Active Living Program on school performance.

Dependent variable: overall school performance	(1) OLS	(2) FE	(3) FE
Treatment * Post	−1.289	−1.090	−1.170
	(0.567)	(0.543)	(0.544)
Adjusted p-value	[0.037]	[0.061]	[0.048]
Treatment (1 = Treatment group)	0.809		
	(0.846)		
Post (1 = 2013/2014, 0 = 2012/2013)	8.589	8.448	2.950
	(0.471)	(0.444)	(3.798)
Age when tests were taken (months)			0.972
			(0.400)
Age squared			−0.002
			(0.001)
Child fixed effects	No	Yes	Yes
Constant	56.860		
	(0.531)		
Observations	2,064	2,028	2,028
R-squared	0.199	0.809	0.812
Number of children in the estimation sample	1,050	1,014	1,014
Mean school performance in estimation sample	61.312	61.222	61.222
Standard deviation	(8.910)	(8.890)	(8.910)
Mean school performance in full sample	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)

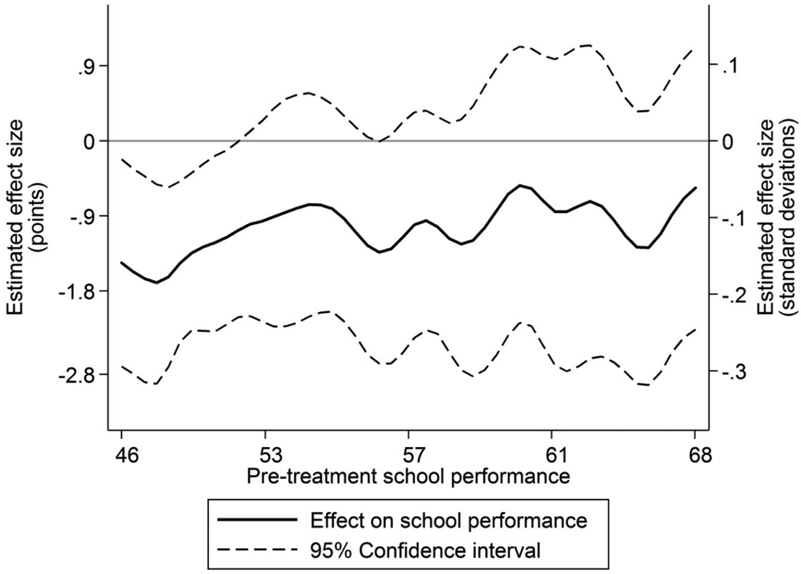
Notes: The table shows the results of three ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

ROBUSTNESS CHECKS

Event Study Analysis

In this section, we investigate whether it is plausible that the negative effects on school performance are indeed due to the Active Living interventions. We therefore conduct an event study analysis in which we repeat our difference-in-differences analysis using interactions of the treatment indicator with three pre-treatment years and the post-treatment year (leaving the 2012/2013 school year as an omitted category).

The results from this analysis show that the difference between the treatment and control group does not change significantly in the pre-treatment years (see Figure 4). This indicates that the pre-treatment trends of the treatment and control groups are relatively parallel. Second, the figure shows that the difference between the treatment and control groups becomes significantly negative in the post-treatment year. This is in line with our previous findings as presented in Table 3. Moreover, there is no sign of a downward trend in the treatment effects in the pre-treatment years, which makes it less likely that the difference between the treatment and control group would also have become negative in the absence of the Active Living Program. Taken together, we conclude that Figure 4 provides strong indications



Notes: The figure shows the smoothed local polynomial graph of the results of ordinary least squares regression analyses across the pre-treatment school performance distribution in 2012/2013. Pre-treatment school performance is a moving window of 200 children. The dependent variable in each regression is overall school performance. The estimated effect size based on difference-in-differences estimations is plotted on the y-axis. Independent variables are a year dummy, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure D4, panel B in the Appendix).[32]
Source: School performance data.

Figure 3. The Effect of the Active Living Program on School Performance, by Pre-Treatment School Performance.

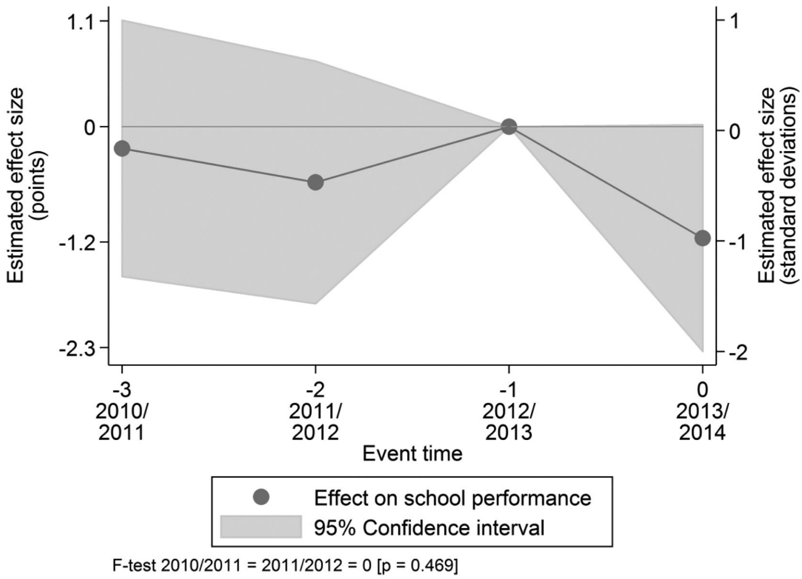
that the parallel trends assumption holds, allowing for a causal interpretation of our estimates of the effects of the Active Living Program.

Repeating the event study analyses for our different specifications of the treatment effect on school performance indicates that the pre-treatment trends in school performance on language tests appear to be parallel as well. However, the estimates of the treatment effect on math performance, as well as those in the right tail of the pre-treatment school performance distribution should be interpreted with caution, given that we only find weak support for the parallel trends assumption in those cases (see Figure D4 in Appendix D).[33]

Additional Robustness Checks

In Appendix B, we show the results of several robustness checks. In the first section of Appendix B, we investigate whether outliers drive our estimates. Due to the

[32] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[33] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.



Notes: The figure shows the results of an ordinary least squares regression analysis. The dependent variable is overall school performance. The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: School performance data.

Figure 4. Event Study Analysis of the Effect of the Active Living Program on School Performance.

relatively small number of schools included in this study and the heterogeneity in the implemented interventions across treatment schools, there is a risk that one particular treatment or control school drives our results. However, the results remain robust across regressions in which we estimate the treatment effects while excluding one different school at a time. We therefore conclude that outlier treatment or control schools do not drive the estimates.

In the second section of Appendix B, we discuss which configuration of design choices (i.e., intervention package) is most strongly associated with the observed negative effect on school performance. Although running the difference-in-differences regressions while *including* only one treatment school at a time (and all control schools) reveals that none of the design choices lead to a positive effect on school performance; the results do not shed more light on the question. The configuration of design choices is so heterogeneous across schools that it is not possible to identify a particular intervention (or combination of interventions) that may drive the results.

In the third section of Appendix B, we analyze whether the negative treatment effect we find is (partially) due to an unexpected increase in school performance in the control group rather than a decrease in school performance in the treatment group. However, difference-in-differences analyses in which we compare control schools to non-Active Living schools lead us to conclude that, even if control schools changed their attitudes or policies toward PA, it does not appear to have had a significant effect on school performance, either in statistical or economic terms.

It is therefore unlikely that unintentional and unobserved changes within control schools bias our estimates of the treatment effect on school performance.

Running the difference-in-differences analyses in which we compare the treatment schools to non-Active Living schools suggests that the estimated effects as presented in Table 3 might be slightly biased upward, and that the more conservative estimate of the average effect of the Active Living Program on school performance is 5.9 percent of a standard deviation.

THE TREATMENT EFFECT ON PHYSICAL ACTIVITY AND EXTERNAL VALIDITY

The robustness checks give strong indications that the negative effect of the Active Living Program on school performance is due to the implemented interventions in the treatment schools. Considering that the Active Living interventions aim to decrease sedentary behavior, we hypothesize that time spent on PA is an important driver of this effect. This also means that in the setting we study, it is still possible to conclude that PA has a positive effect on school performance if the Active Living Program (unexpectedly) has a negative effect on time spent on PA. In order to analyze this, we use the accelerometer data.[34]

We identify the effect of the Active Living Program on time spent on PA by estimating the following model:

$$PA_{ist} = \beta_0 + \beta_1 \left(Treat_{is} * Post_t\right) + \beta_2 Post_t + \boldsymbol{X}'_{\boldsymbol{ist}}\beta + \gamma_{3is} + \varepsilon_{3ist}. \quad (3)$$

The *Treat* and *Post* variables, as well as child fixed effects γ, are identical to those in equation (1). $\boldsymbol{X'}$ is a vector of variables, including the time children wore the accelerometer per day and several variables indicating the weather conditions during each PA measurement week (i.e., temperature, hours of sunshine, and rainfall). To allow for nonlinear relationships between these variables and time spent on PA, we also control for accelerometer wearing time squared, temperature squared, and hours of sunshine squared.[35] Considering the Active Living interventions implemented, we expect to find that the difference-in-differences estimator β_1 is positive and significantly different from zero.

Table 4 reports the results of the difference-in-differences analyses on time spent on PA during school time and leisure time. Based on these results, we conclude that the Active Living Program significantly increases time spent on PA during school hours by 9.30 minutes per school day. This effect is equal to 0.34 standard deviations. There appears to be a small crowding-out effect of leisure-time PA, as the Active Living Program decreases time spent on PA during leisure time by 3.8 minutes per day (0.08 standard deviations), but this effect is not significantly different from zero.

Additional analyses indicate that the Active Living Program has a positive effect on time spent on PA during school hours of approximately 10 minutes per day across the pre-treatment school performance distribution. The Active Living Program does not have a significant effect on leisure-time PA in almost all parts of the school performance distribution (see Figure D6 in Appendix D).[36]

[34] Additional analyses indicate that wearing an accelerometer does not in itself have any statistically or economically significant effects on school performance.

[35] Because the study design for the Active Living Program required that the pre- and post-treatment PA measurements were scheduled around the same time for each school, age is approximately constant in the PA data once we control for year and child fixed effects.

[36] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 4. Difference-in-differences estimations of the effect of the Active Living Program on time spent on physical activity during school time and leisure time.

	(1)	(2) School time	(3)	(4)	(5) Leisure time	(6)
Dependent variable: time spent on PA (min./day)	OLS	FE	FE	OLS	FE	FE
Treatment * Post	25.337	22.544	9.298	−2.247	−5.638	−3.814
	(8.552)	(8.611)	(3.283)	(7.836)	(7.126)	(3.853)
Adjusted p-value	[0.009]	[0.017]	[0.016]	[0.778]	[0.439]	[0.344]
Treatment (1 = Treatment group)	−14.990			1.414		
	(5.243)			(5.293)		
Post (1 = 2013/2014, 0 = 2012/2013)	−15.175	−13.062	−13.963	−18.765	−19.374	−24.136
	(5.872)	(6.179)	(3.182)	(4.426)	(4.494)	(5.104)
Wearing time of accelerometer (min./day)			0.790			0.412
			(0.305)			(0.191)
Wearing time squared			−0.001			−0.000
			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			−0.269			0.184
			(0.069)			(0.146)
Temperature squared			0.001			−0.000
			(0.000)			(0.001)

Table 4. Continued.

	(1)	(2) School time	(3)	(4)	(5) Leisure time	(6)
Dependent variable: time spent on PA (min./day)	OLS	FE	FE	OLS	FE	FE
Amount of sunshine during PA measurement week (hours/day)			8.483			10.339
			(2.258)			(4.660)
Sunshine squared			−0.771			−0.853
			(0.268)			(0.520)
Rain during PA measurement week (1 = yes)			1.441			−2.597
			(1.820)			(4.654)
Child fixed effects	No	Yes	Yes	No	Yes	Yes
Constant	100.280			200.506		
	(4.111)			(3.049)		
Observations	1,355	1,204	1,072	1,301	1,120	1,018
R-squared	0.058	0.111	0.503	0.041	0.137	0.435
Number of children in the estimation sample	753	602	536	741	560	509
Mean time spent on PA in estimation sample	91.510	91.973	93.500	191.819	192.564	192.978
Standard deviation	(26.989)	(26.756)	(25.183)	(49.545)	(48.696)	(46.767)
Mean time spent on PA in full sample	91.510	91.510	91.510	191.819	191.819	191.819
Standard deviation	(26.989)	(26.989)	(26.989)	(49.545)	(49.545)	(49.545)

Notes: The table shows the results of six ordinary least squares regression analyses. The dependent variable in columns (1) through (3) is time spent on PA during school time (min./day). The dependent variable in columns (4) through (6) is time spent on PA during leisure time (min./day). Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: PA data.

Instrumental Variables Estimation

The estimated effects of the Active Living Program on time spent on PA can help interpret the size of the effect of PA on school performance. The causal effect of time spent on PA on school performance can be calculated using the Wald estimator, by dividing the estimated treatment effect on school performance (i.e., the reduced form, captured by α_1) by the estimated treatment effect on time spent on PA (i.e., the first stage, captured by β_1). Under the assumption that the Active Living Program only affects school performance through school-time PA, we would conclude that an extra 30 minutes spent on PA during school hours reduces school performance by 1.76 score points (18.9 percent of a standard deviation).[37] This result needs to be interpreted with caution; although the effect of the Active Living Program on school-time PA is statistically significant (t-statistic of 2.89), it does not meet the criterion of a strong instrument.

Note that the assumption of instrument exogeneity appears to be strong as it is conceivable that the Active Living Program also affected other potential confounders (e.g., increased leisure-time PA and improved health and well-being from adjusted playgrounds or safer roads around school). However, as these examples indicate, it is implausible that such confounders are *negatively* related to school performance. Therefore, even if the exclusion restriction would be violated by these confounders, they cannot explain the *negative* effect on school performance; controlling for such variables would make the negative effect on school performance even stronger. Considering the robust negative effects of the Active Living Program on school performance, we therefore conclude that time spent on PA during school hours appears to have a negative effect on school performance in the context of this study.

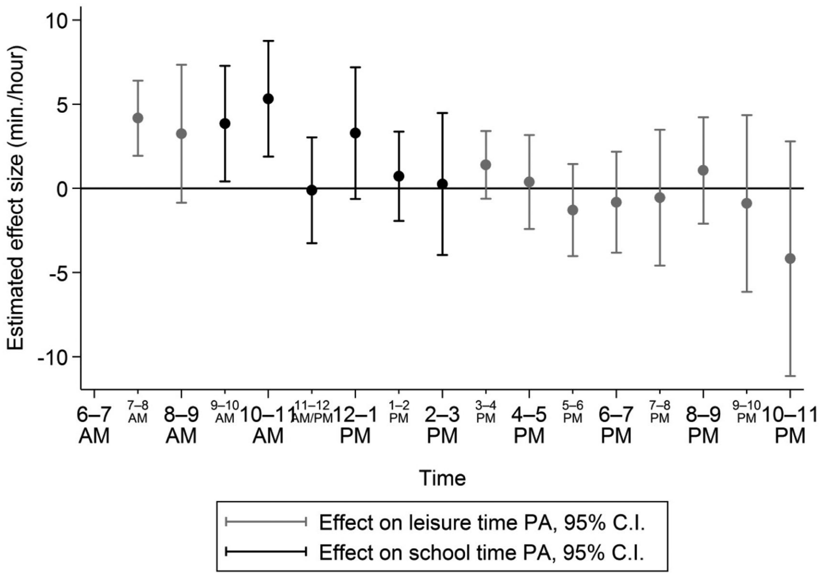
External Validity

Appendix C discusses the external validity of our findings. We use data from the 2015 Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) to show that exercising or practicing a sport before going to school is negatively correlated with 15-year-old students' performance within 60 countries and economies across the world. Moreover, in all countries, this correlation is even more negative in the full sample than within the sample of children of low-educated parents. Evidently, these analyses are not causal in nature and should be interpreted with caution. However, the similarity between our "naïve" pre-treatment OLS estimates and the Wald estimator indicates that the bias of the OLS might be small, and that, if anything, the OLS estimate is slightly biased toward zero. The results in Appendix C therefore seem to indicate that it is possible that increasing PA in the mornings of school days can also have negative effects on school performance in other high-income countries such as the United States, Sweden, or Japan, as well as in middle-income countries such as Brazil, Colombia, or Tunisia. Additionally, they suggest that our main results are generalizable outside of low-SES regions.

[37] When using results from Table B4 (column 6) and Table 4 (column 3), a back-of-the-envelope calculation of the Wald estimator is −0.059 (−0.546 score points/9.298 minutes), implying that for each minute increase in school-time PA, overall school performance decreases by 0.059 points (0.63 percent of a standard deviation). All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

MECHANISMS

There may be several mechanisms through which time spent on PA during school hours may affect school performance. Before analyzing potential mechanisms, we first analyze in more detail at what time of the day the effect on school-time PA is strongest. Figure 5 shows the estimated treatment effects on time spent on PA



Notes: The figure shows the results of 16 ordinary least squares regression analyses. The dependent variable in each regression is time spent on PA on school days (excluding Wednesdays). The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axis. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The effect of the Active Living Program on time spent on PA between 6:00 a.m. and 7:00 a.m. could not be estimated because the number of observations is too small.
Source: PA data.

Figure 5. The Effect of the Active Living Program on Time Spent on PA, by Hour on School Days (Excluding Wednesdays).

by hour on school days.[38] When interpreting this figure, it is important to note that children typically commute to school between 8:00 a.m. and 9:00 a.m., and that recess time typically takes place between 10:00 a.m. and 11:00 a.m. (around 15 minutes) and between 12:00 p.m. to 1:00 p.m. (around one hour), meaning that most instruction time takes place between 9:00 a.m. and 10:00 a.m., 11:00 a.m. and 12:00 p.m., and 1:00 p.m. and 3:00 p.m.

It appears that the Active Living Program affects time spent on PA between 7:00 a.m. and 1:00 p.m., but not in the afternoons or evenings of school days. The figure highlights that the Active Living Program does not appear to crowd out leisure-time PA on school days; if anything, it may have increased active transportation to school (positive effect between 8:00 a.m. and 9:00 a.m.). Second, the

[38] We exclude Wednesdays because on these days, children have a different time schedule than on the other weekdays.

figure highlights that some of the additional time spent on PA during school hours has taken place during instruction time (positive effect between 9:00 a.m. and 10:00 a.m.).

Figure 5 suggests that PA increased in the mornings of schooldays. This insight can help us pinpoint which mechanisms may underlie the negative effects of the Active Living Program on school performance. The effects on school performance may be negative because increased time spent on PA crowds out other beneficial activities in the mornings of school days. The effect could also be negative due to behavioral changes *after* school. Because PA after school is not affected, this may be due to a substitution of sedentary activities, meaning that children substitute more cognitively stimulating activities (e.g., playing music or doing homework) for less cognitively stimulating ones. This section discusses these potential mechanisms in turn.

Lateness Due to Active Transportation

The most important (sedentary) activity that affects school performance and takes place in the mornings of schooldays is instruction time. Although the interventions do not affect the curriculum and instruction time, one could argue that increased active transportation to school (positive effect on PA between 8:00 a.m. and 9:00 a.m.) may increase lateness or attendance, which would decrease instruction time received by the child. Although we cannot exclude the possibility that children in treatment schools arrived later at school than usual, this is unlikely because schools are usually very strict about the time students need to be at school. Moreover, it is unlikely that active transportation instead of inactive transportation delayed children by more than five minutes because the average distance from home to primary school is approximately 700 meters (Statistics Netherlands, 2017).

Restlessness During Instruction Time

Another way in which the Active Living Program might negatively affect instruction time is through decreased quality of the instruction time. One can argue that spending time on PA, and active play in particular, can be fun, social, and tiring. This, in turn, might make children more distracted, restless, or chatty during instruction time, especially at the moments shortly after the activity. The children's restlessness may render instruction time less effective, even if only a few children are affected (due to negative spillover effects). The fact that some of the additional time spent on PA during school hours took place during instruction time indicates the possibility that, due to the Active Living Program, children in treatment schools may have been more restless during instruction time (compared to the control group).

Moreover, we can investigate this potential mechanism in more detail across cohorts of sixth graders. As part of the Onderwijs Monitor Limburg project, every year since 2008, sixth graders have been asked to complete a questionnaire. Questions related to the mechanisms of restlessness during instruction time are: "Give yourself a report card score. You can give yourself the following scores: 5 = insufficient, 6 = sufficient, 7 = more than sufficient, 8 = good. How good are you at: . . . Being calm and quiet when the teacher wants me to" (this question was asked of sixth graders from 2011/2012 through 2014/2015), "Staying in my seat when the teacher wants me to" and "Being studious and not being a pest" (these two questions were asked of sixth graders from 2011/2012 until 2013/2014). Although these personal characteristics are self-assessed by the children, they do give some insight into the attitudes and behaviors during instruction time. Children who score higher on these variables on average have better school performance (see Tables D4 to D6 in Appendix D).[39]

Given that these questions are only asked of sixth graders, we do not have pre- and post-treatment measurements for the same individuals. We can, however, run difference-in-differences analyses by comparing pre- and post-treatment cohorts within treatment and control schools. The key assumption for causal identification with this strategy is that the difference between the characteristics in the 2012/2013, 2013/2014, and 2014/2015 cohorts in treatment schools would have been similar to the difference between these cohorts in control schools if the Active Living Program had not taken place. We estimate the following model:

$$Restless_{isc} = \theta_0 + \theta_1 \left(Treat_{is} * Post'_c\right) + \theta_2 Treat_{is} + \theta_3 Post'_c + \varepsilon_{4isc}. \qquad (4)$$

In this equation, i indexes individual pupil, s indexes school, and c indexes cohorts of sixth graders. *Restless* stands for any of the three "restlessness variables," which are treated as continuous variables. The *Post'* variable is a cohort indicator that takes the value of one for children who were enrolled in grade 6 in 2013/2014. The difference-in-differences estimator then takes on the value of one for sixth graders in treatment schools in 2013/2014 (i.e., the treatment year of the Active Living Program), and it takes on the value of zero for sixth graders in treatment schools in 2012/2013, as well as for sixth graders in control schools in any given school year. The *Treat* variable is identical to those in equations (1) through (3).

Figure 6 reveals that, due to the Active Living Program, children in treatment schools report being less able to be calm and quiet when the teacher wants them to. This relationship is statistically significant and appears to persist across time, although it is noisier after two years and no longer statistically significant. The results in Figure 6 are also still significantly different from zero when we apply a Bonferroni method to account for multiple hypotheses testing. The size of the effect is relatively large and can explain around 52 percent of the estimated effect of time spent on PA during school hours on school performance.[40] Figures D7 and D8 in Appendix D[41] show that the Active Living Program has no significant effect on the variables "Good at: Staying in my seat" and "Good at: Being studious, not a pest."

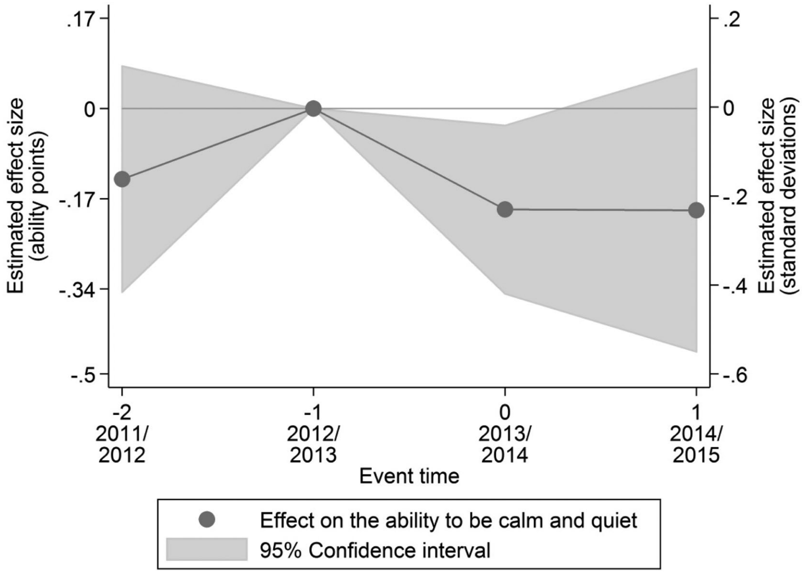
Crowding Out of Other Cognitively Stimulating Morning Activities

Children's cognitively stimulating morning activities that take place outside of instruction time (i.e., before going to school and in recess time) could include doing

[39] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

[40] This is calculated as follows. Using results from Table B4 (column 6) and Table 4 (column 3), we calculate that our most conservative Wald estimator of the causal effect of time spent on PA during school hours on school performance is −0.059 (−0.546 score points/9.298 minutes). In other words, each additional minute spent on PA during school time decreases overall school performance by 0.059 points (0.63 percent of a standard deviation). Using the results from Figure 6 and Table 4 (column 3), we also conclude that each additional minute spent on PA during school hours decreases children's ability to remain calm and quiet by 0.020 points (−0.190 ability points/9.298 minutes). In Table D5 we saw that each additional point in the ability to remain calm and quiet is generally related to school performances that are 1.535 points higher. We would therefore expect that the 0.020 point reduction in the ability to remain calm and quiet, caused by a one-minute increase in school-time PA, would decrease school performance by 0.031 score points (−0.020 ability points * 1.535). This is equal to around 52 percent of the causal effect of time spent on PA during school hours on school performance (−0.031 score points/−0.059 score points).

[41] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.



Notes: The figure shows results of an ordinary least squares regression analysis. The dependent variable is children's self-assessed ability to be calm and quiet when the teacher wants them to (on a scale of 5 to 8). The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are cohort fixed effects and a treatment dummy. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: Onderwijs Monitor Limburg questionnaire.

Figure 6. The Effect of the Active Living Program on Children's Ability to be Calm and Quiet When the Teacher Wants Them To.

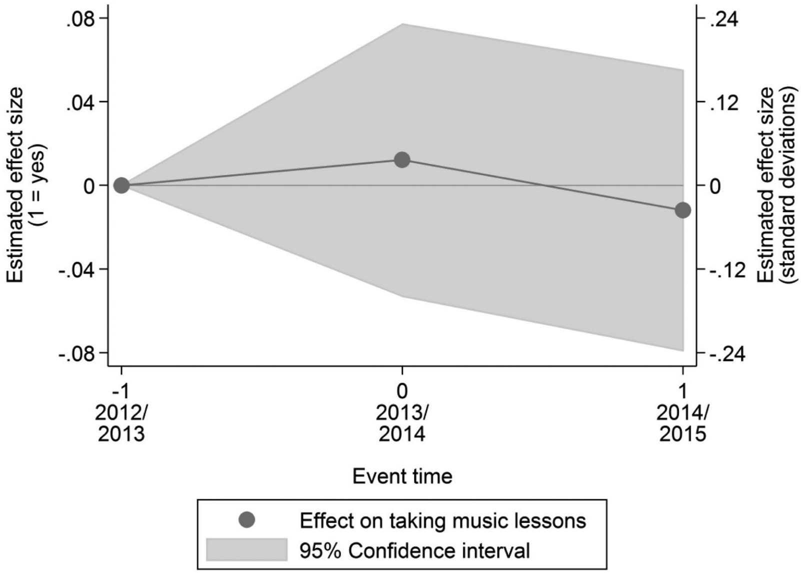
homework or talking with parents and friends. However, PA can be a socializing activity as well, making it unlikely to be less cognitively stimulating than talking with parents or friends. Although it is unlikely that children do homework during recess, it remains possible that active transportation to school crowded out doing homework before going to school. Due to data limitations, we unfortunately cannot test this hypothesis.

Substitution of Sedentary Afternoon Activities

The negative effect on school performance could be explained by another mechanism: increased time spent on PA during school hours increases fatigue after school hours, leading children to substitute sedentary cognitively stimulating activities in leisure time (e.g., playing music, reading, or doing homework) for other sedentary but less cognitively stimulating activities (e.g., watching television).[42] Again, we can analyze these potential mechanisms across cohorts of sixth graders.

The only question in the Onderwijs Monitor Limburg questionnaire related to cognitively stimulating afternoon activities is: "Which of the following activities do you

[42] Because we do not find that the Active Living Program increased time spent on PA during leisure time, we assume that children substituted one sedentary activity for another during leisure time.



Notes: The figure shows the results of an ordinary least squares regression analysis. The dependent variable is whether the child takes music lessons (1 = yes). The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are cohort fixed effects and a treatment dummy. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: Onderwijs Monitor Limburg questionnaire.

Figure 7. The Effect of the Active Living Program on Taking Music Lessons.

do in a club or a group? Music lessons . . . "[43] This question was asked to sixth graders from 2012/2013 through 2014/2015.[44] Considering the negative treatment effect on school performance, we hypothesize that the Active Living Program may have decreased the probability of taking music lessons (assuming that music lessons are substituted by less cognitively stimulating activities in leisure time). Taking music lessons is positively related to school performance (see Table D5 in Appendix D).[45]

Figure 7 shows that we cannot conclude that the Active Living Program has any effect on the probability of taking music lessons; the estimated treatment effects are not significant, either in statistical or economic terms. Due to data limitations, we also cannot rule out the possibility that the Active Living Program had a (negative) effect on time spent doing homework, reading, or participating in other cognitively stimulating post-school activities.

[43] Other activities that could be ticked are: football, tennis, chess, arts and crafts, hockey, scouts, singing, dancing/ballet, swimming, and horseback riding. We do not find evidence that the sports activities in this list are significantly affected by the Active Living Program, which is not surprising considering that we do not find effects on PA in the afternoons of schooldays. The only sedentary activities in this list (i.e., "playing chess" and "arts and crafts") are not significantly affected by the Active Living Program.
[44] The variable provides the lower bound of the number of children who play instruments, because some children may play musical instruments without taking lessons.
[45] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Unobservable Potential Mechanisms

Taken together, our analyses show that PA may negatively affect test scores through a reduction in the ability to be calm and quiet in class (i.e., increased restlessness). Another way to interpret this result is that the Active Living Program may have altered the interpretation of the school setting and its values and expectations. It is possible that the colorful play equipment and fun games led children to have a more playful and less "academic" perception of the school setting (picked up by our finding of a reduced ability to remain calm and quiet in class), which in turn negatively affects the average school performance.

The restlessness mechanism can explain roughly half of the effect of the Active Living Program on school performance. Due to data limitations, other potential mechanisms that may account for the remaining unexplained part of the negative treatment effect on school performance remain uncertain. For instance, increased time spent on PA may have crowded out investments in potentially beneficial activities during leisure time, such as reading, studying, or social activities. Moreover, it remains possible that increased time spent on PA also crowded out potentially harmful activities such as smoking or watching television, which would attenuate the negative effect on school performance.

CONCLUSIONS

This paper investigates whether encouraging children to become more physically active in their everyday life affects their primary school performance. Our estimations are based on a field quasi-experiment called the Active Living Program, which aimed to increase physical activity and decrease sedentary behavior among 8- to 12-year-olds living in low-SES areas in the Netherlands. The interventions focus on active transportation and active play at school and during leisure time. Results from difference-in-differences analyses reveal that the Active Living Program significantly decreases school performance by at least 5.9 percent of a standard deviation in the treatment year. This result is robust across a variety of specifications, but the negative effect is strongest among the worst-performing students. The results of event study analyses and several robustness checks provide strong indications that the negative effect is indeed due to the Active Living interventions causing a significant decrease in the treatment group's school performance.

Besides its effects on school performance, the Active Living Program causes a significant increase in time spent on physical activity during school time of 0.34 standard deviations (9.30 minutes per school day). The effect remains robust across the pre-treatment school performance distribution. The Active Living Program does not significantly affect leisure-time physical activity. We conclude that time spent on physical activity during school hours appears to have a negative (short-term) effect on children's school performance in the context of this study. This effect seems largely driven by children's reduced ability to remain calm and quiet when the teacher wants them to.

We can put our estimates of the treatment effects into perspective by comparing them to the estimated effect sizes on teenagers' academic achievement of other well-known interventions. Anderson (2008) reviews the treatment effects of Abercedarian (ABC), the Perry Preschool, and Early Training Projects (ETP) in the United States.[46]

[46] Beginning as early as 1962, these programs targeted disadvantaged African-Americans in North Carolina, Michigan, and Tennessee. Through random assignment, treated children in each experiment received several years of preschool education (with intensity differing across programs). Interventions continued until the children began regular schooling (Anderson, 2008).

Although these interventions may not be directly comparable to the Active Living Program due to different target groups and contexts, they do provide potential benchmarks for our estimates. If we simply compute an average of the treatment effects weighted by sample size,[47] the overall impact of the three programs on teen outcomes is about 0.28 standard deviations. This is almost five times as large as our estimates of the effect of the Active Living Program on school performance.

Our results shed new light on the effects of the encouragement of physical activity among children and adolescents. To our knowledge, this study is the first to analyze the causal effect of encouraging informal physical activity on educational outcomes. Previous studies have shown a positive relationship between formal physical activity (i.e., participating in sports or physical education) and school performance, but based on our findings, we conclude that these results appear not to be generalizable to informal physical activity such as active transportation or active play. One reason for this may be that formal physical activity is more effective for fostering skills that are beneficial for school performance, such as confidence, discipline, or competitiveness, than informal physical activity. Moreover, because exercising is generally more intense than being active in everyday life, formal physical activity may yield higher returns in terms of health and well-being. If children usually reallocate study hours (or other cognitively stimulating tasks such as paying attention during instruction hours) to physical activity, the combination of the absence of the development of sports-related skills and the potentially lower health benefits could explain why the encouragement of informal physical activity has a negative effect on school performance.

Additionally, our results indicate that the encouragement of informal physical activity may increase inequality in school performance. The Active Living Program focuses on children living in low-SES areas, where (pre-treatment) school performance is lower. This means that the Active Living Program negatively affects school performance in the worse-performing regions of the Netherlands, and that within these regions, the negative effect is strongest among the worst-performing students.

Although our results may not be generalizable to interventions that encourage formal physical activity or that focus on adults, we do find suggestive evidence of the possibility that increasing adolescents' physical activity in the mornings of school days can also have negative effects on school performance in other high- or middle-income countries and outside of low-SES regions. This means that this study has important implications for policymakers, educators, and researchers around the world who aim to increase physical activity among children and adolescents. First, this study highlights that they cannot assume that every type of physical activity intervention will only have positive externalities if it is effective in increasing time spent on physical activity. To make an informed decision about whether to increase the time children spend on physical activity, policymakers and educators must weigh the benefits for health and well-being of the child relative to the potential cost of decreased educational performance. In this process, they implicitly or explicitly make a choice as to what extent education outcomes and health outcomes are important for children. This means that despite the existing literature that points toward a positive relationship between physical activity and educational outcomes, policymakers must remain alert when implementing interventions, and carefully monitor all the (assumed) benefits and potential costs. Another important takeaway is that the costs and benefits of physical activity may not be the same for all children. The worst-performing students, who need their study and instruction time the most, may also suffer the most if other cognitively stimulating activities

[47] This calculation is identical to Deming (2009) and is based on Table 3 of Anderson (2008).

are crowded out or if the quality of their instruction time decreases because they are tired or distracted from being active. To prevent increased educational inequalities, targeted interventions may be needed that increase low-performing students' physical activity without further decreasing their educational performance. Finally, before resources are allocated to increase time spent on physical activity among children and adolescents, we first need to improve our understanding of the causal effects of such interventions on short-term and long-term educational outcomes. Future research needs to focus on which forms of physical activity do not crowd out desirable activities.

BART H. H. GOLSTEYN is a Professor in the Department of Economics at Maastricht University, School of Business and Economics, P. O. Box 616, 6200MD Maastricht, the Netherlands (e-mail: b.golsteyn@maastrichtuniversity.nl).

MARIA W. J. JANSEN is a Professor in the Department of Health Services Research at the Care and Public Health Research Institute (CAPHRI) at Maastricht University, P. O. Box 616, 6200MD Maastricht, the Netherlands, and the Academic Collaborative Centre for Public Health Limburg, Public Health Services (e-mail: maria.jansen@ggdzl.nl).

DAVE H. H. VAN KANN is a Senior Researcher in the School of Sport Studies at Fontys University of Applied Sciences, P.O. Box 347, 5600 AH Eindhoven, and in the Department of Health Promotion at Maastricht University (e-mail: d.vankann@fontys.nl).

ANNELORE M. C. VERHAGEN (corresponding author) is a PhD candidate in the Department of Economics at Maastricht University, School of Business and Economics, P. O. Box 616, 6200MD Maastricht, the Netherlands (e-mail: amcverhagen@gmail.com).

ACKNOWLEDGMENTS

We are grateful to all schools and children participating in this study and to all Maastricht University colleagues involved in the data collection. Previous versions of the paper were presented at the 21st IZA Summer School in Labour Economics, the Maastricht University Workshop in Economics, the Applied Economics Lunch Seminar at the Paris School of Economics, the Learning and Work seminar at Maastricht University, the Netherlands Economists Day 2017, the 29th annual conference of the European Association of Labour Economists, the 31st Annual Conference of the European Society for Population Economics, and the Human Enhancement and Learning conference. We thank these conference and seminar participants, as well as Eric Bonsang, Lex Borghans, Stefano DellaVigna, Thomas Dohmen, Ilyana Kuziemko, Michael Lechner, and Bas Ter Weel, for valuable comments and suggestions. This study was funded by the Netherlands Organization for Scientific Research (VIDI grant 452-16-006) and the Netherlands Organization for Health Research and Development (ZonMW), Project Number 200130003.

REFERENCES

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association, 103, 1481–1495.

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Barron, J., Ewing, B., & Waddell, G. (2000). The effects of high school athletic participation on education and labor market outcomes. Review of Economics and Statistics, 82, 409–421.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology, 28, 169–181.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust difference-in-differences estimates? The Quarterly Journal of Economics, 119, 249–275.

Cabane, C., Hille, A., & Lechner, M. (2016). Mozart or Pelé? The effects of adolescents' participation in music and sports. Labour Economics, 41, 90–103.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics, 90, 414–427.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. The Journal of Human Resources, 50, 318–372.

Carver, A., Timperio, A., & Crawford, D. (2013). Parental chauffeurs: What drives their transport choice? Journal of Transport Geography, 26, 72–77.

Cawley, J., Frisvold, D., & Meyerhoefer, C. (2013). The impact of physical education on obesity among elementary school children. Journal of Health Economics, 32, 743–755.

Cawley, J., Meyerhoefer, C., & Newhouse, D. (2007). The impact of state physical education requirements on youth physical activity and overweight. Health Economics, 16, 1287–1301.

Centers for Disease Control and Prevention. (2011). Strategies to prevent obesity and other chronic diseases: The CDC guide to strategies to increase physical activity in the community. Atlanta, GA: U.S. Department of Health and Human Services.

Choi, L., Liu, Z., Matthews, C. W., & Buchowski, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. Medicine, and Science in Sports, and Exercise, 43, 357–364.

Colcolmbe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive functioning of older adults: A meta-analytic study. Psychological Science, 14, 125–130.

Deming, D. (2009). Early childhood intervention and life-cycle skills development: Evidence from Head Start. American Economic Journal: Applied Economics, 1, 111–134.

Dills, A. K., Morgan, H. N., & Rotthoff, K. W. (2011). Recess, physical education, and elementary school student outcomes. Economics of Education Review, 30, 889–900.

Donald, D. G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. The Review of Economics and Statistics, 89, 221–223.

Erickson, K. I., Hillman, C. H., & Kramer, A. F. (2015). Physical activity, brain, and cognition. Current Opinion in Behavioral Sciences, 4, 27–32.

European Commission. (2014). EU action plan on childhood obesity 2014/2020.

Evenson, K., Catellier, D., Gill, K., Ondrak, K., & McMurray, R. (2008). Calibration of two objective measures of physical activity for children. Journal of Sports Sciences, 26, 1557–1565.

Felez-Nobrega, M., Hillman, C. H., Cirera, E., & Puig-Ribera, A. (2017). The association of context-specific sitting time and physical activity intensity to working memory capacity and academic achievement in young adults. The European Journal of Public Health, 27, 741–746.

Felfe, C., Lechner, M., & Steinmayr, A. (2016). Sports and child development. PLoS ONE, 11, 1–23.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. Nature Reviews Neuroscience, 9, 58–65.

Hume, C., Timperio, A., Salmon, J., Carver, A., Giles-Corti, B., & Crawford, D. (2009). Walking and cycling to school: Predictors of increases among children and adolescents. American Journal of Preventive Medicine, 36, 195–200.

Lipscomb, S. (2007). Secondary school extracurricular involvement and academic achievement: A fixed effects approach. Economics of Education Review, 26, 463–472.

Lizandra, J., Devís-Devís, J., Pérez-Gimeno, E., Valencia-Peris, A., & Peiró-Velert, C. (2016). Does sedentary behavior predict academic performance in adolescents or the other way round? A longitudinal path analysis. PLoS ONE, 11, 1–13.

Pfeifer, C., & Cornelissen, T. (2010). The impact of participation in sports on educational attainment—New evidence from Germany. Economics of Education Review, 29, 94–103.

Pont, K., Ziviani, J., Wadley, D., Bennett, S., & Abbott, R. (2009). Environmental correlates of children's active transportation: A systematic literature review. Health & Place, 15, 849–862.

Rees, D. I., & Sabia, J. J. (2010). Sports participation and academic performance: Evidence from the National Longitudinal Study of Adolescent Health. Economics of Education Review, 29, 751–759.

Rich, C., Griffiths, L., & Dezateux, C. (2012). Seasonal variation in accelerometer-determined sedentary behaviour and physical activity in children: A review. International Journal of Behavioral Nutrition and Physical Activity, 9 1–8.

Singh, A., Uijtdewilligen, L., Twisk, J. W. R., van Mechelen, W., & Chinapaw, M. J. M. (2012). Physical activity and performance at school: A systematic review of the literature including a methodological quality assessment. Archives of Pediatrics and Adolescent Medicine, 166, 49–55.

Statistics Netherlands. (2013). Statistics on population density via statline.cbs.nl.

Statistics Netherlands. (2014). Statline: nabijheid voorzieningen; afstand locatie, wijk- en buurtcijfers 2013. Voorburg/Heerlen: Statistics Netherlands.

Statistics Netherlands. (2017). Statline. http://statline.cbs.nl/.

Stevenson, B. (2010). Beyond the classroom: Using Title IX to measure the return to high school sports. Review of Economics and Statistics, 92, 284–301.

Strong, W., Malina, R., Blimkie, C., Daniels, S., Dishman, R., & Gutin, B. (2005). Evidence-based physical activity for school-age youth. The Journal of Pediatrics, 146, 732–737.

van den Berg, V., Salasi, E., de Groot, R. H. M., Jolles, J., Chinapaw, M. J. M., & Singh, A. S. (2016). Physical activity in the school setting: Cognitive performance is not affected by three different types of acute exercise. Frontiers in Psychology, 7, 1–9.

van Kann, D. H. H., Jansen, M. W. J., de Vries, S. I., & Kremers, S. P. J. (2015). Active Living: development and quasi-experimental evaluation of a school-centered physical activity intervention for primary school children. BMC Public Health, 15, 1–10.

van Kann, D. H. H., Kremers, S. P. J., de Vries, N. K., de Vries, S. I., & Jansen, M. W. J. (2016). The effect of a school-centered multicomponent intervention on daily physical activity and sedentary behavior in primary school children: The Active Living study. Preventive Medicine, 89, 64–69.

World Health Organization. (2010). Global recommendations on physical activity for health. Geneva, Switzerland: WHO Press.

APPENDIX A: CONTROL VARIABLES AND SUMMARY STATISTICS

Control Variables

In the school performance data, age is registered as the average age (in months) on the day the tests were taken. In the PA data, age is registered as age (in months) during the PA measurement week. Both data sets include an indicator variable with a value of one if the child is a boy. We create an indicator variable with a value of one if the child was enrolled in grade 5 during the pre-treatment year and zero if in that school year, the child was enrolled in grade 4. The accelerometer wearing time is registered in minutes. Our weather indicators are mean temperature, hours of sunshine, and rainfall (millimeters) between 6:00 a.m. and 11:00 p.m. during the PA measurement week. Rainfall is transformed into an indicator variable with a value of one if more than zero millimeters of rain fell. The weather indicators are based on information from the weather station of the Royal Dutch Meteorological Institute (KNMI) located at the Maastricht-Aachen airport and are calculated for every day of the measurement week.

Summary Statistics

Table A1 lists descriptive statistics of our full sample, i.e., of children in the target group (fourth and fifth graders in 2012/2013) who are enrolled in one of the 21 Active Living schools, as well as of their peers in non-Active Living schools. First, the table shows that the children in Active Living schools account for 17 percent of our full sample. It also highlights that Active Living schools are located in low-SES areas; average school performance is significantly lower in Active Living schools compared to the other (non-Active Living) schools in our data. Children in Active Living schools and non-Active Living schools are comparable with respect to age, gender, and grade in which they were enrolled in 2012/2013.

Table A2 lists descriptive statistics for the estimation samples of our main results (i.e., including child fixed effects and a vector of control variables as specified earlier in the section on Empirical Strategy). In our estimation samples, we have school performance data from 1,014 children, school-time PA data from 536 children, and leisure-time PA data from 509 children. When the school performance data are merged with the PA data, school-time PA data remain for 422 children and leisure-time PA data for 398 children.

In the school performance data, 52 percent of the sample is enrolled in a treatment school. This share is slightly larger in the PA data (56 percent), which indicates that there are relatively more children in control schools for whom we do not have (valid) PA data.

On average, children took five tests per year (i.e., the number of tests on which the school performance measure is based). The average school performance in the pre-treatment year is 57.28, with a standard deviation of 7.90. Because the subject scores are standardized across school years and children's performance generally improves as they get older, scores increase over time, to an average of 65.16 with a standard deviation of 8.04 in the post-treatment year.

The children in the estimation sample were, on average, 10.5 years old (126 months) when they took their pre-treatment tests, and 11.5 years old (138 months) when they took tests in the post-treatment year. Because we exclude tests that were taken during the PA measurement weeks, the age of the children in the school performance data is, on average, two months lower than in the PA data.

Table A1. Descriptive statistics for Active Living schools and non-Active Living schools.

	Full sample		Non–Active Living schools		Active Living schools			
	mean	(sd)	mean	(sd)	mean	(sd)	Diff.	P-value
Child enrolled in an Active Living school (1 = yes)	0.17	(0.38)	0.00	(0.00)	1.00	(0.00)	−1.00	0.000
Overall school performance	62.28	(9.30)	62.41	(9.33)	61.65	(9.11)	0.76	0.001
Language performance	61.62	(9.25)	61.74	(9.29)	61.04	(9.01)	0.70	0.002
Math performance	60.96	(8.88)	61.10	(8.88)	60.30	(8.86)	0.80	0.000
Age when tests were taken (months)	132.39	(10.25)	132.37	(10.25)	132.46	(10.26)	−0.09	0.718
Gender (1 = boy)	0.49	(0.50)	0.49	(0.50)	0.48	(0.50)	0.02	0.189
Cohort (1 = 5th graders, 0 = 4th graders)	0.61	(0.49)	0.61	(0.49)	0.59	(0.49)	0.02	0.129
Number of children in the estimation sample	6,776		5,671		1,105			
Number of schools	174		153		21			

Notes: Descriptive statistics are based on the pre-treatment year (2012/2013) and the post-treatment year (2013/2014) combined.
Source: School performance data.

Table A2. Descriptive statistics for the estimation samples.

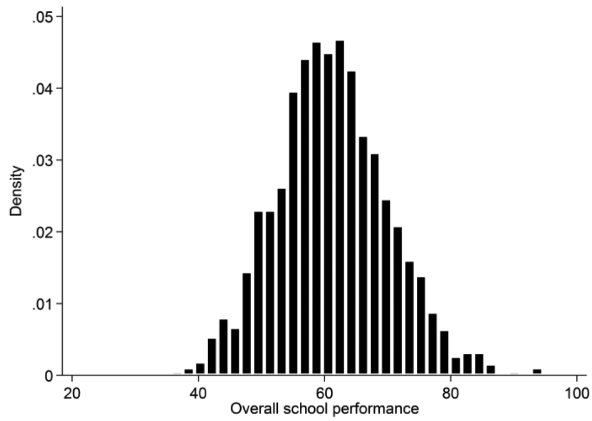
	Pre-treatment year		Post-treatment year			
	mean	(sd)	mean	(sd)	Diff.	P-value
Panel A. School performance data						
Treatment (1 = Treatment group)	0.52	(0.50)	0.52	(0.50)	0.00	1.000
Number of tests taken	5.00	(0.93)	4.98	(0.77)	−0.02	0.515
Overall school performance	57.28	(7.90)	65.16	(8.04)	7.88	0.000
Language performance	56.99	(7.89)	64.29	(8.09)	7.31	0.000
Math performance	56.05	(7.99)	63.81	(7.64)	7.76	0.000
Age when tests were taken (months)	126.10	(8.35)	137.91	(8.21)	11.81	0.000
Gender (1 = boy)	0.48	(0.50)	0.48	(0.50)	0.00	1.000
Cohort (1 = 5th graders, 0 = 4th graders)	0.57	(0.49)	0.57	(0.49)	0.00	1.000
Number of children with school performance data	1,014		1,014			
Panel B. PA data						
Treatment (1 = Treatment group)	0.56	(0.50)	0.56	(0.50)	0.00	1.000
Number of days the accelerometer was worn	4.10	(1.15)	4.15	(1.46)	0.05	0.547
Time spent on school-time PA	93.36	(24.74)	93.64	(25.64)	0.28	0.854
Time spent on leisure-time PA	203.50	(47.13)	182.46	(43.99)	−21.04	0.000
Wearing time of accelerometer during school (min./day)	309.23	(32.18)	323.73	(31.67)	14.51	0.000
Wearing time of accelerometer during leisure (min./day)	547.89	71.55	527.24	69.71	−20.65	0.000
Age during PA measurement (months)	127.75	(8.22)	139.70	(8.23)	11.96	0.000
Gender (1 = boy)	0.44	(0.50)	0.44	(0.50)	0.00	1.000
Cohort (1 = 5th graders, 0 = 4th graders)	0.45	(0.50)	0.45	(0.50)	0.00	1.000
Outside temperature during PA measurement (Celsius)	13.07	(5.58)	17.04	(3.45)	3.97	0.000
Amount of sunshine during PA measurement week (hours/day)	2.62	(1.59)	5.20	(2.26)	2.58	0.000
Rain during PA measurement week (1 = yes)	0.46	(0.50)	0.47	(0.50)	0.01	0.669
Number of children with school-time PA data	536		536			
Number of children with leisure-time PA data	509		509			

Table A2. Continued.

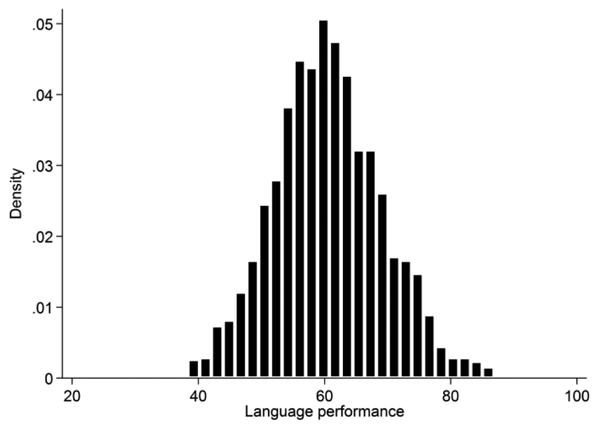
	Pre-treatment year		Post-treatment year			
	mean	(sd)	mean	(sd)	Diff.	P-value
Panel C. Merged data						
Treatment (1 = Treatment group)	0.60	(0.49)	0.60	(0.49)	0.00	1.000
Overall school performance	57.67	(7.94)	65.58	(8.14)	7.91	0.000
Language performance	57.38	(7.92)	64.73	(8.23)	7.35	0.000
Math performance	56.24	(7.76)	64.09	(7.48)	7.85	0.000
Time spent on school-time PA	91.84	(23.65)	92.12	(25.02)	0.28	0.867
Time spent on leisure-time PA	202.65	(47.78)	182.24	(44.81)	−20.40	0.000
Wearing time of accelerometer during school time (min./day)	308.75	(31.47)	325.62	(32.40)	16.87	0.000
Wearing time of accelerometer during leisure time (min./day)	546.82	(70.43)	529.76	(69.61)	−17.06	0.001
Age when tests were taken (months)	124.89	(8.27)	136.74	(8.11)	11.85	0.000
Age during PA measurement (months)	128.36	(8.34)	140.31	(8.34)	11.95	0.000
Gender (1 = boy)	0.42	(0.49)	0.42	(0.49)	0.00	1.000
Cohort (1 = 5th graders, 0 = 4th graders)	0.52	(0.50)	0.52	(0.50)	0.00	1.000
Outside temperature during PA measurement (Celsius)	12.46	(5.72)	16.80	(3.36)	4.34	0.000
Amount of sunshine during PA measurement week (hours/day)	2.59	(1.62)	4.96	(2.39)	2.37	0.000
Rain during PA measurement week (1 = yes)	0.44	(0.50)	0.51	(0.50)	0.07	0.039
Number of children with school performance and school-time PA data	422		422			
Number of children with school performance and leisure-time PA data	398		398			

Notes: Descriptive statistics for the school performance data are based on the estimation sample of the analyses presented in column 3 of Table 3 in the main text. Descriptive statistics for the PA data are based on the estimation sample of the analyses presented in columns 3 and 6 of Table 4 in the main text. Descriptive statistics for the merged data are based on the estimation sample of the analyses presented in Figure 17 in Appendix D.

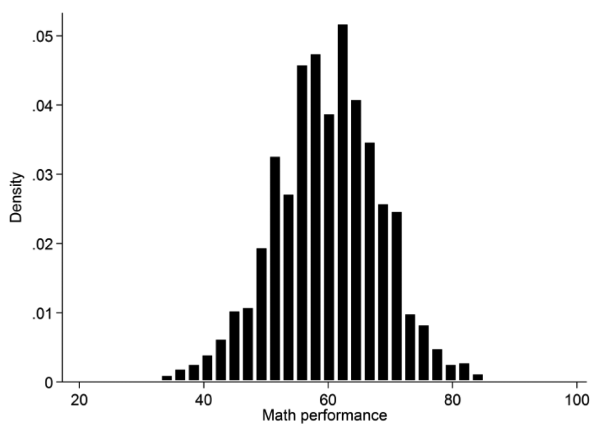
Panel A. Overall school performance



Panel B. Language

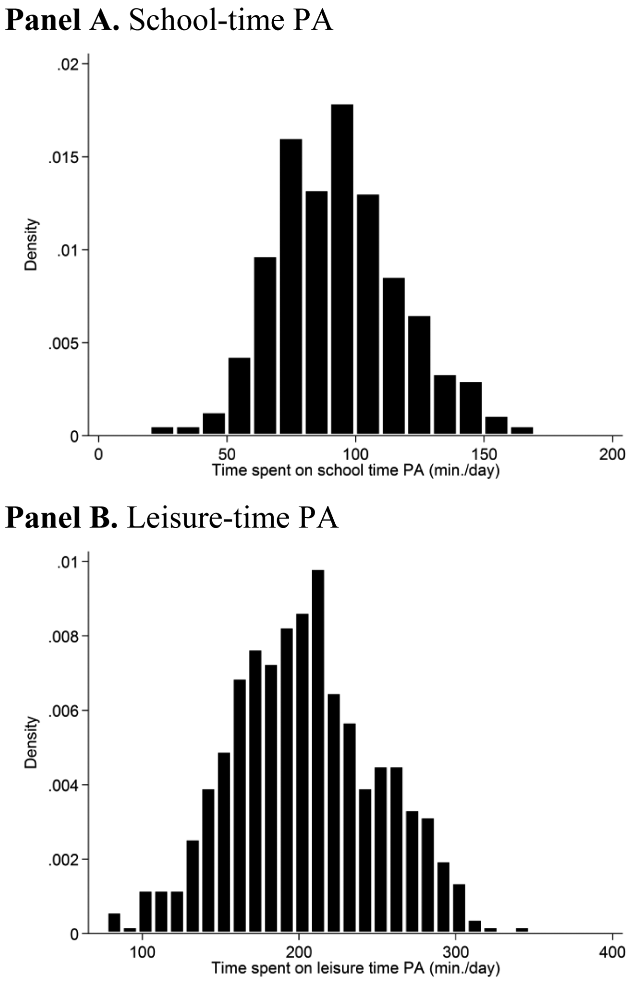


Panel C. Math



Notes: Calculations are based on the estimation sample of the analysis presented in column 3 of Table 3 in the main text.
Source: School performance data.

Figure A1. Distribution of School Performance in the Estimation Sample.



Notes: Calculations are based on the estimation sample of the analyses presented in columns 3 and 6 of Table 4 in the main text.
Source: PA data.

Figure A2. Distribution of Time Spent on Physical Activity in the Estimation Sample.

In each data set, a little less than half the sample is male. The slight overrepresentation of the 5th-grade cohort in the school performance data shows that there are more fourth graders than fifth graders who did not take any tests.

The weather conditions were slightly better during the PA measurements in the treatment year compared to the PA measurements in the pre-treatment year: outside temperatures were slightly higher and there were more hours of sunshine.

Figures A1 and A2 indicate that the pre-treatment measures of school performance and time spent on PA are approximately normally distributed.

APPENDIX B: ROBUSTNESS CHECKS

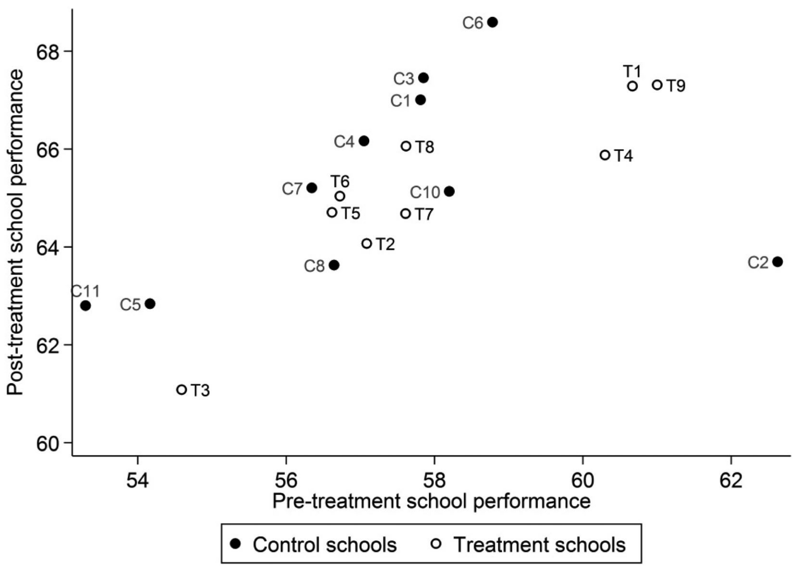
Checking for Outliers

In our main text, we show that the pre-treatment trends in school performance appear to be parallel between the treatment and control groups. Here, we verify whether our results may be driven by an outlier school. This is of particular concern due to the relatively small number of schools included in this study and because the treatment is not uniform across the treatment schools. Figure B1 indicates that it is unlikely that the results are driven by any particular treatment school, i.e., any particular intervention package.

We next run regressions excluding one different school each time (see Tables B1 and B2). Considering that the results remain robust across these regressions, we conclude that outlier treatment or control schools do not drive the estimates.

Configuration of Design Choices

As Table 1 in the main text indicates, there is considerable heterogeneity in the implemented interventions across treatment schools. Having such heterogeneity in implementation is, in some respects, a strength because it likely mimics a real-world implementation. However, this also makes it harder to interpret the treatment effects. Although Figure B1 and Table B1 both indicate that the treatment effects on school performance are relatively similar between treatment schools, the question



Notes: The figure is based on the estimation sample of the analysis presented in Table 3, column 3. The numbers of the treatment schools correspond to those in Table 1 in the main text. The results in Tables 7 and 8 show that the difference-in-differences estimations of the effect of the Active Living Program on overall school performance remain robust when one of the Active Living schools is excluded from the analysis.
Source: School performance data.

Figure B1. Average Pre- and Post-Treatment School Performance, by Active Living School.

Table B1. Difference-in-differences estimations of the effect of the Active Living Program on school performance—one-by-one deletion of treatment schools.

Dependent variable: overall school performance	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
Treatment * Post	−1.052	−1.161	−1.075	−1.096	−1.325	−1.314	−1.111	−1.329	−1.083
	(0.543)	(0.556)	(0.562)	(0.547)	(0.561)	(0.555)	(0.553)	(0.550)	(0.550)
Adjusted p-value	[0.073]	[0.056]	[0.076]	[0.065]	[0.033]	[0.033]	[0.064]	[0.030]	[0.069]
Post (1 = 2013/2014, 0 = 2012/2013)	2.632	2.006	4.334	3.174	2.517	3.016	1.514	3.418	2.373
	(3.996)	(3.972)	(4.240)	(3.853)	(3.907)	(3.848)	(3.490)	(3.821)	(3.824)
Age when tests were taken (months)	1.045	1.085	0.881	0.938	0.949	0.954	0.980	0.993	0.966
	(0.416)	(0.411)	(0.452)	(0.407)	(0.448)	(0.408)	(0.404)	(0.413)	(0.401)
Age squared	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.001	−0.002	−0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,904	1,958	1,896	1,978	1,826	1,874	1,918	1,898	1,942
R-squared	0.813	0.813	0.815	0.814	0.807	0.810	0.823	0.811	0.814
Number of children in the estimation sample	952	979	948	989	913	937	959	949	971
Mean school performance in estimation sample	61.042	61.245	61.457	61.174	61.267	61.249	61.226	61.179	61.092
Standard deviation	(8.911)	(8.921)	(8.814)	(8.898)	(8.847)	(8.888)	(8.947)	(8.863)	(8.980)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of nine ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tn = Treatment school (number), which is excluded from the analysis. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

Table B2. Difference-in-differences estimations of the effect of the Active Living Program on school performance—one-by-one deletion of control schools.

Dependent variable: overall school performance	(1) Excl. C1	(2) Excl. C2	(3) Excl. C3	(4) Excl. C4	(5) Excl. C5	(6) Excl. C6	(7) Excl. C7	(8) Excl. C8	(9) Excl. C9	(10) Excl. C10	(11) Excl. C11
Treatment * Post	−1.142	−1.294	−0.890	−1.145	−1.158	−1.044	−1.122	−1.373	−1.170	−1.515	−1.051
	(0.589)	(0.538)	(0.508)	(0.562)	(0.575)	(0.564)	(0.621)	(0.564)	(0.544)	(0.492)	(0.571)
Adjusted p-value	[0.073]	[0.030]	[0.102]	[0.061]	[0.064]	[0.085]	[0.092]	[0.029]	[0.048]	[0.008]	[0.087]
Post (1 = 2013/2014, 0 = 2012/2013)	3.024	2.298	2.197	3.324	2.859	4.497	2.571	6.034	2.950	1.782	2.987
	(3.919)	(3.806)	(4.471)	(4.077)	(3.814)	(3.773)	(3.938)	(3.234)	(3.798)	(4.168)	(3.741)
Age when tests were taken (months)	1.089	0.997	0.865	0.983	1.013	0.923	0.927	0.686	0.972	1.217	0.955
	(0.412)	(0.409)	(0.428)	(0.410)	(0.407)	(0.409)	(0.417)	(0.344)	(0.400)	(0.408)	(0.400)
Age squared	−0.002	−0.002	−0.001	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,944	2,012	1,892	1,988	1,958	1,950	1,874	1,892	2,028	1,868	1,932
R-squared	0.809	0.819	0.814	0.810	0.809	0.809	0.804	0.816	0.812	0.820	0.807
Number of children in the estimation sample	972	1,006	946	994	979	975	937	946	1,014	934	966
Mean school performance in estimation sample	61.170	61.206	61.133	61.214	61.319	61.123	61.258	61.300	61.222	61.183	61.379
Standard deviation	(8.883)	(8.914)	(8.712)	(8.901)	(8.855)	(8.895)	(8.952)	(8.918)	(8.890)	(8.982)	(8.805)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of 11 ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Cn = Control school (number), which is excluded from the analysis. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

remains whether the intervention packages that cause the strongest treatment effects have something in common. To answer this question, we run the difference-in-differences regressions while each time including only one treatment school (and all control schools), instead of excluding them one by one as we did in Table B1. This allows us to identify which configuration of design choices (i.e., intervention package) is most strongly associated with the observed negative effect on school performance. However, this strategy reduces the number of observations drastically: we go from having test scores of 1,014 children to having scores from only 500 children, which makes the interpretation of insignificant effects problematic due to a potential lack of power.

Based on the results shown in Table B3, we first conclude that none of the design choices lead to a positive effect on school performance. The strongest negative treatment effect is found in treatment schools 1, 4, and 9. The negative treatment effect on school performance is smallest in treatment schools 5, 6, and 8. However, the configuration of design choices is so heterogeneous across schools that it is not possible to identify a particular intervention (or combination of interventions) that may drive these results.

Comparisons to Non-Active Living Schools

We cannot completely rule out the possibility that control schools changed certain attitudes or policies toward PA. However, we can analyze whether there were changes in school performance in control schools relative to all other schools in our data that did not participate in the Active Living Program as a treatment or control group (i.e., the non-Active Living schools). By doing so, we can verify whether the negative treatment effects we find are indeed due to a decrease in school performance in the treatment group, and not due to an (unexpected) increase in school performance in the control group.

When we compare the school performance of children in the treatment group to those of children in non-Active Living schools, the difference-in-differences model we estimate is similar to equation (1) in the main text, with the exception that the *Treat* variable now has the value of zero for children in the non-Active Living schools and is set to "missing" for children in the control schools.

When we compare the children in the control schools to those in the non-Active Living schools, the difference-in-differences model we estimate is as follows:

$$Score_{ist} = \alpha_{11} + \alpha_{12}\left(Control_{is} * Post_t\right) + \alpha_{13}Post_t + \alpha_{14}age_{ist} + \alpha_{15}age_{ist}^2 + \gamma_{4is} + \varepsilon_{4ist}. \tag{B.1}$$

Control is an indicator variable with a value of one for children in control schools and zero for children in non-Active Living schools. The other variables are identical to those in equation (1) in the main text. Although pre-treatment differences in school performance between control schools and non-Active Living schools exist because all Active Living schools are located in relatively deprived areas (as shown in Table D2 in Appendix D), these differences should not change significantly over time. In other words, α_{12} should not be significantly different from zero, for otherwise it indicates that the control group may (unexpectedly) be affected by the Active Living Program, which would complicate the interpretation of any effects of the Active Living Program we find. However, if we indeed cannot reject the hypothesis that the difference-in-differences estimator α_{12} is equal to zero, we have strong reasons to believe that the effects of the Active Living Program are the result of the Active Living interventions in the treatment schools. This conclusion becomes even

Table B3. Difference-in-differences estimations of the effect of Active Living on school performance—one-by-one inclusion of treatment schools.

Dependent variable: Overall school performance	(1) Incl. T1	(2) Incl. T2	(3) Incl. T3	(4) Incl. T4	(5) Incl. T5	(6) Incl. T6	(7) Incl. T7	(8) Incl. T8	(9) Incl. T9
Treatment * Post	−2.167	−1.347	−1.591	−2.612	−0.507	−0.321	−1.647	−0.085	−2.194
	(0.529)	(0.527)	(0.486)	(0.423)	(0.498)	(0.513)	(0.536)	(0.469)	(0.466)
Adjusted p-value	[0.005]	[0.038]	[0.014]	[0.000]	[0.343]	[0.551]	[0.018]	[0.862]	[0.002]
Post (1 = 2013/2014, 0 = 2012/2013)	−0.161	1.937	−0.107	0.267	1.402	0.622	2.264	−0.065	0.525
	(4.609)	(4.616)	(4.299)	(4.737)	(4.574)	(4.670)	(5.319)	(4.624)	(4.760)
Age when tests were taken (months)	1.028	0.872	1.069	1.038	1.049	1.049	1.094	1.031	1.058
	(0.603)	(0.627)	(0.575)	(0.649)	(0.548)	(0.630)	(0.650)	(0.614)	(0.649)
Age squared	−0.001	−0.001	−0.001	−0.001	−0.002	−0.001	−0.002	−0.001	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Child fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,094	1,040	1,102	1,020	1,172	1,124	1,080	1,100	1,056
R-squared	0.829	0.827	0.825	0.827	0.836	0.832	0.811	0.831	0.826
Number of children in the estimation sample	547	520	551	510	586	562	540	550	528
Mean school performance in estimation sample	61.412	61.050	60.695	61.182	61.037	61.056	61.090	61.173	61.334
Standard deviation	8.978	8.967	9.126	9.014	9.075	9.017	8.919	9.064	8.857
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of nine ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tn = Treatment school (number), which is included in the analysis. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

stronger if the point estimate of the effect of the Active Living Program is similar to our main result (i.e., negative) when we compare the treatment group to the non-Active Living schools.

The results in Table B4 show that the school performance of children in the control group slightly increases due to the Active Living Program (0.07 standard deviations), but as expected, this effect is not significantly different from zero. Therefore, even if control schools changed their attitudes or policies towards PA, this does not appear to have had a significant effect on school performance, either in statistical or economic terms. It is therefore unlikely that unintentional and unobserved changes within control schools bias our estimates of the treatment effect on school performance.

The effect of the Active Living Program on the treatment group's school performance remains negative and significantly different from zero at the 10 percent level in the main specification, although it is now smaller in size (-0.06 standard deviations). This smaller effect may be explained by the positive, though insignificant, "treatment" effect on the control group's school performance. Although caution is needed when interpreting results that are not statistically significant, this suggests that the estimated effects as presented in Table 3 (main text) might be slightly biased upward, and that the more conservative estimate of the average effect of the Active Living Program on school performance is 5.9 percent of a standard deviation.[48]

[48] We also ran a regression of school performance on participating in the Active Living Program as a treatment school in which we control for baseline school performance instead of including the *Treat*Post* interaction. The estimated treatment effect using this specification is -0.06 standard deviations, significantly different from zero at the 10 percent level.

Table B4. Difference-in-differences estimations of the effect of the Active Living Program on school performance—Children in Active Living schools compared to children in non-Active Living schools.

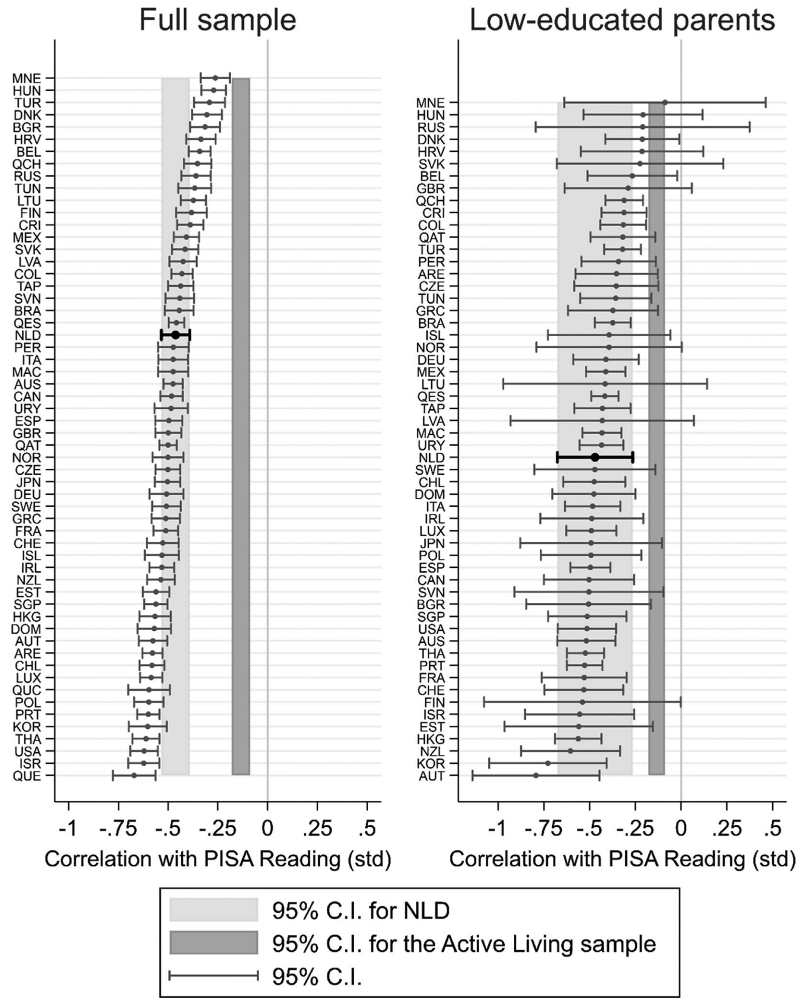
	(1)	(2) Control group	(3)	(4)	(5) Treatment group	(6)
Dependent variable: overall school performance	OLS	FE	FE	OLS	FE	FE
Control * Post	0.342	0.765	0.664			
	(0.501)	(0.453)	(0.466)			
Adjusted p-value	[0.496]	[0.094]	[0.157]			
Treatment * Post				−0.090	−0.325	−0.546
				(0.880)	(0.334)	(0.301)
Adjusted p-value				[0.919]	[0.331]	[0.072]
Control (1 = Control group, 0 = non–Active Living)	−1.056					
	(0.606)					
Treatment (1 = Treatment group, 0 = non–Active Living)				−0.246		
				(0.715)		
Post (1 = 2013/2014, 0 = 2012/2013)	8.247	7.683	−1.351	8.247	7.683	−1.080
	(0.199)	(0.133)	(1.465)	(0.199)	(0.133)	(1.487)
Age when tests were taken (months)			1.566			1.571
			(0.190)			(0.185)
Age squared			−0.003			−0.003
			(0.001)			(0.001)
Child fixed effects	No	Yes	Yes	No	Yes	Yes
Constant	57.916			57.916		
	(0.313)			(0.313)		
Observations	11,361	10,350	10,348	11,480	10,438	10,436
R-squared	0.197	0.788	0.801	0.195	0.784	0.797
Number of children in the estimation sample	6,186	5,175	5,174	6,261	5,219	5,218
Mean school performance in estimation sample	62.311	61.665	61.667	62.367	61.687	61.689
Standard deviation	(9.313)	(9.173)	(9.173)	(9.309)	(9.143)	(9.143)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: The table shows the results of six ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

APPENDIX C: EXTERNAL VALIDITY

Although the internal validity of our findings appears to be strong, it is uncertain to what extent our findings are externally valid (generalizable) beyond the setting of this study, i.e., 8- to 12-year-olds living in low-SES regions in the Netherlands. Does encouraging informal PA also have a negative effect on school performance (1) in high-SES regions in the Netherlands, (2) in low- or high-SES regions in other countries, and (3) among children in different age categories? Future research should address these questions. In this paper, we can only provide suggestive evidence. For instance, data from the 2015 OECD Programme for International Student Assessment (PISA) show that, within 60 countries and economies across the world, exercising or practicing a sport before going to school is negatively correlated with 15-year-olds' student performance (see Figure C1). Although the PA is not necessarily informal in nature, the time at which it takes place is comparable to the time of day at which the Active Living Program has the strongest positive effect on PA (i.e., in the mornings of schooldays). Moreover, the correlation is negative in all countries, and it is even more negative in the full sample than from within the sample of children from low-educated parents.[49] Evidently, these analyses are not causal in nature. However, the similarity between our "naïve" pre-treatment OLS estimates (see Table 2 in the main text) and the Wald estimator (see the section on Instrumental Variables Estimation in the main text) indicate that the OLS might be biased toward zero. Figure C1 therefore suggests that we may expect that increasing PA in the mornings of school days can also have negative effects on school performance in other high-income countries such as the United States, Sweden, or Japan, as well as in middle-income countries such as Brazil, Colombia, or Tunisia. Additionally, it suggests that the results are generalizable outside of low-SES regions (within and outside of the Netherlands), and among children in different age categories (i.e., 8- to 12-year-olds as well as 15-year-olds). However, it remains unlikely that we can generalize our findings to Physical Education classes, high school athletics, or PA after school hours including participation in sports clubs. It is possible that the difference in the effect that formal and informal PA have on school performance is related to the potential mechanism we find. Children may find it harder to remain calm and quiet shortly after their activities than they do shortly before potential activities. More research is needed to identify the mechanisms for the effects of different types of PA on school performance.

[49] The results for children of low-educated parents are similar to the results for children in the bottom quartile of the SES indicator as derived by the OECD. Moreover, the results are qualitatively robust for controlling for gender and age, and for different measures of student performance (i.e., PISA Mathematics and PISA Science).



Notes: The figure shows the within-country correlation between students' answer to the question "On the most recent day you attended school, did you do any of the following before going to school? Exercise or practice a sport" (ST076Q11NA) and PISA Reading (standardized within the country-level estimation samples). "95 percent C.I. for the Active Living sample" refers to the pre-treatment correlation between spending 30 minutes on physical activity during school time and performance on language tests. Low-educated parents are those whose highest obtained degree (combined) is at most ISCED level 2 (lower secondary education/second stage of basic education).
Source: PISA 2015.

Figure C1. Correlation Between Physical Activity Before Going to School and Reading Performance.

APPENDIX D: ADDITIONAL TABLES AND FIGURES

Textbox D1. Additional information on the selection of treatment and control schools.

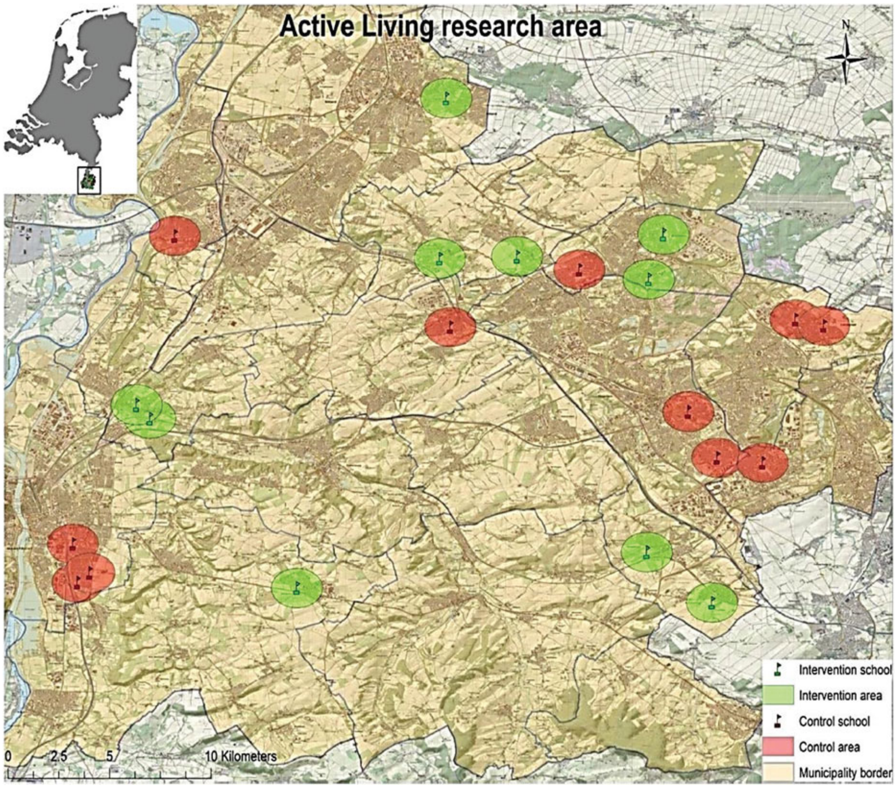
First, municipal support for participation was essential because costs for environmental adaptations in public spaces (i.e., school surroundings) had to be covered by municipalities. In November 2011, municipal development plans of all 19 Southern-Limburg municipalities were checked to see if they contained formal references to the themes of “youth” and “overweight prevention.” Municipalities were excluded from the list of potential treatment areas if they did not state these topics in their development plans because this implied that no budget was reserved for environmental adaptations for these purposes. In total, 12 out of 19 municipalities had formulated targets either for youth or overweight prevention or both. Municipal health officers in these 12 municipalities were visited and informed about the project and conditions for participation. This resulted in six municipalities wanting to become involved as an intervention area. In the six municipalities that agreed to participate, four school corporations were identified, three of which agreed to recruit schools falling under their responsibility. To participate in the Active Living project as a treatment school, schools were checked for eligibility according to predefined inclusion criteria: (1) located in a deprived area; (2) at least 25 students enrolled in grades 4 and 5; (3) no plans to merge with another school or plans to relocate in the upcoming 3 years; and (4) willing to actively participate and to form an “Active Living” working group at school. Within the municipalities and school corporations that consented, 37 primary schools were identified in deprived areas. Municipalities and corporations were asked to recommend schools that were most eligible to participate from their perspective. We visited 13 eligible schools and informed them about the project and conditions for participation; 10 schools (76.9 percent) agreed to participate as a treatment school (see Figure D2).

After recruiting treatment schools, we defined the potential control areas as the municipalities that were excluded in the process of treatment school selection, i.e. the seven municipalities without development plans including references to the themes of youth and/or overweight prevention, and the six municipalities that were not interested in participating as a treatment area or not able to cover the expenditures of environmental adaptations. From these municipalities, control schools were chosen based on matching on the level of SES and degree of urbanization. Control schools were offered Public Health Services (PHS) support to implement effective elements of the project after the end of the effectiveness study.

Note: This textbox is based on Van Kann et al. (2015).

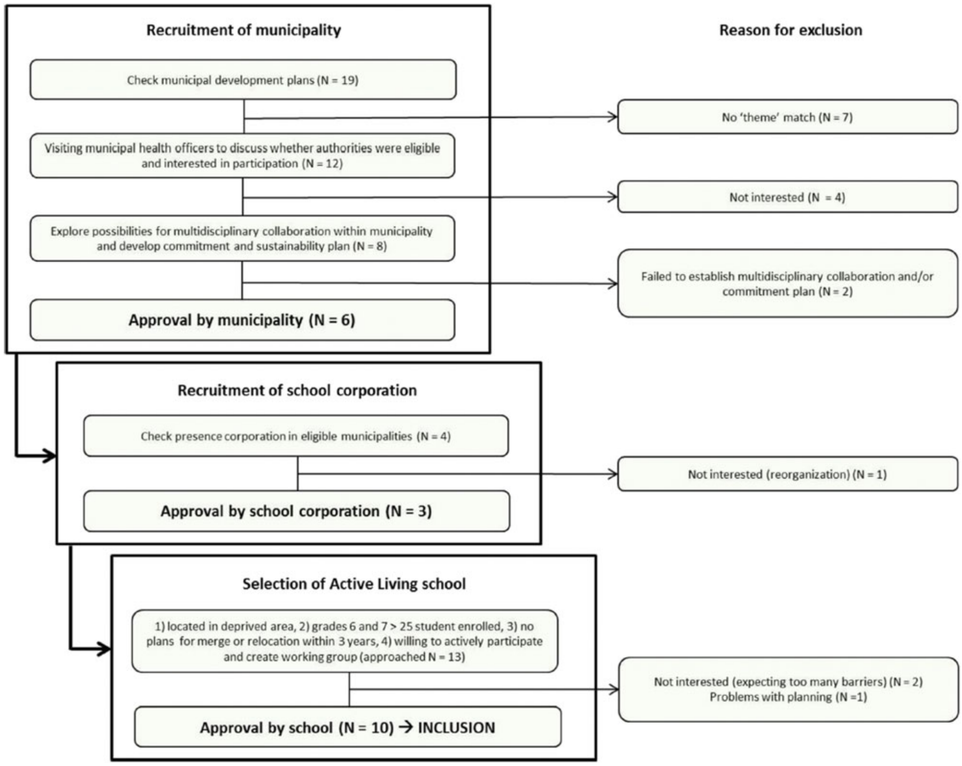
Source: Pictures taken by the research team.

Picture D1. Pictures of Implemented Interventions.



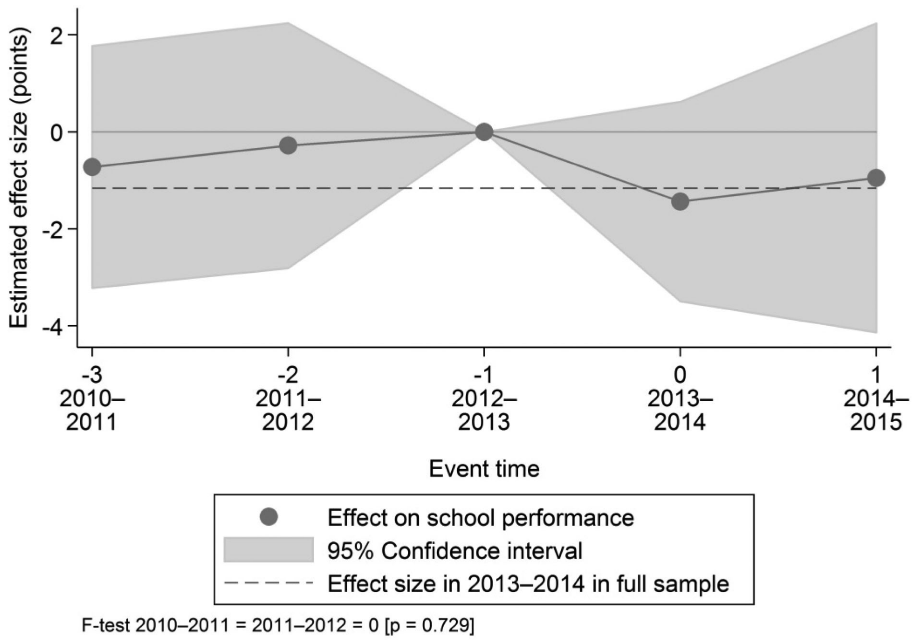
Source: Van Kann et al. (2015).

Figure D1. Research Area—Southern-Limburg Region, The Netherlands.



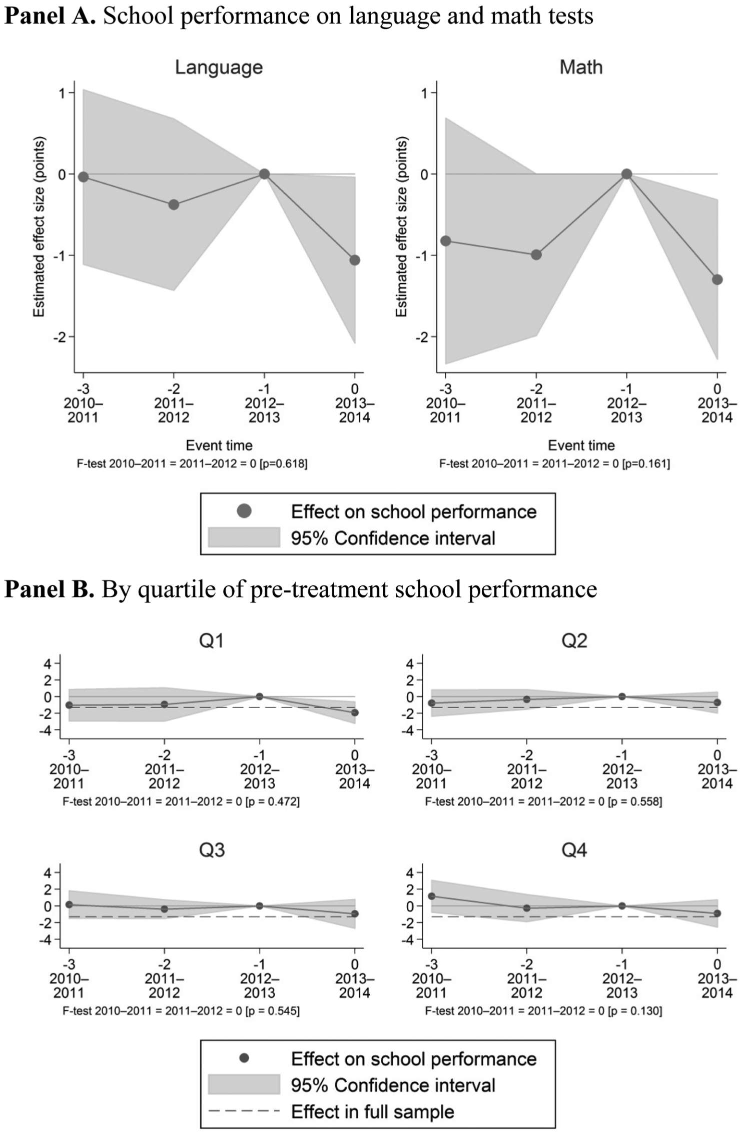
Source: Van Kann et al. (2015).

Figure D2. Recruitment of Treatment Schools.



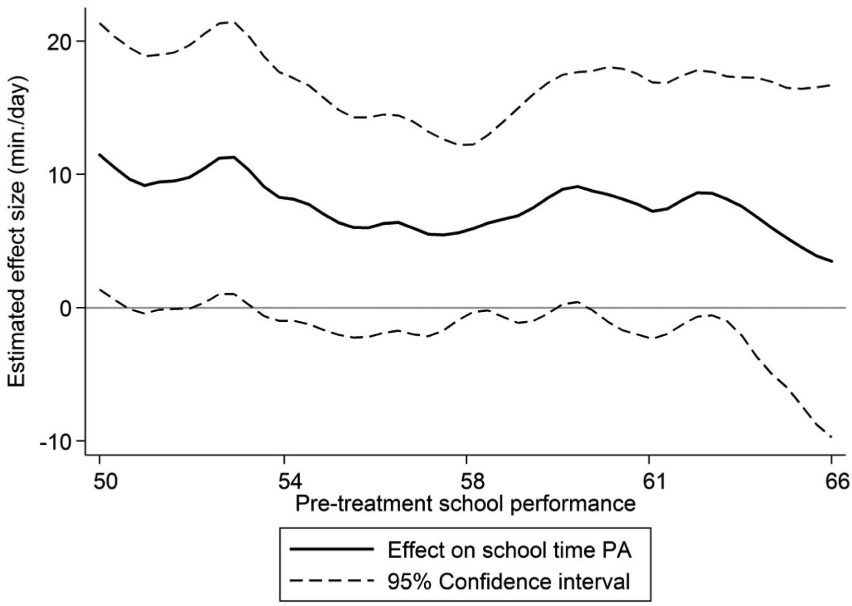
Notes: The figure shows the results of an ordinary least squares regression analysis including only children who were enrolled in grade 4 in the pre-treatment year 2012/2013. The dependent variable is overall school performance. The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: School performance data.

Figure D3. The Long-Run Effects of the Active Living Program on School Performance.



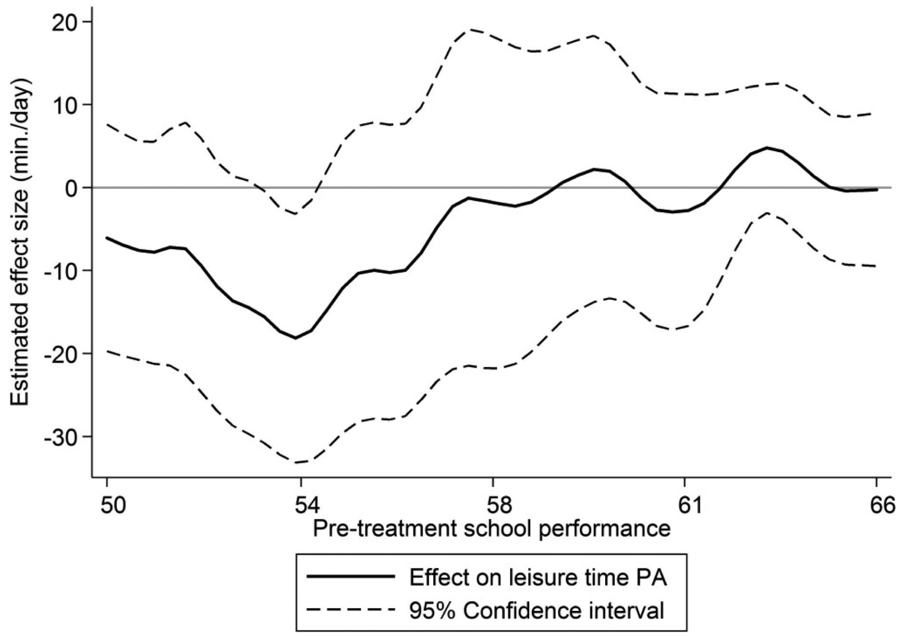
Notes: The figure shows the results of six ordinary least squares regression analyses. Panel A is estimated on the full sample, and panel B is estimated by quartile of the pre-treatment school performance distribution in 2012/2013. The dependent variable in each regression is overall school performance. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are year fixed effects, age, age squared, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. Panel A depicts the 90 percent confidence interval, because the results in Table D3 show that the treatment effects in the treatment year are significant at the 90 percent level.
Source: School performance data.

Figure D4. Event Study Analyses of the Effect of the Active Living Program on Language and Math Tests, and Across the Pre-Treatment School Performance Distribution.



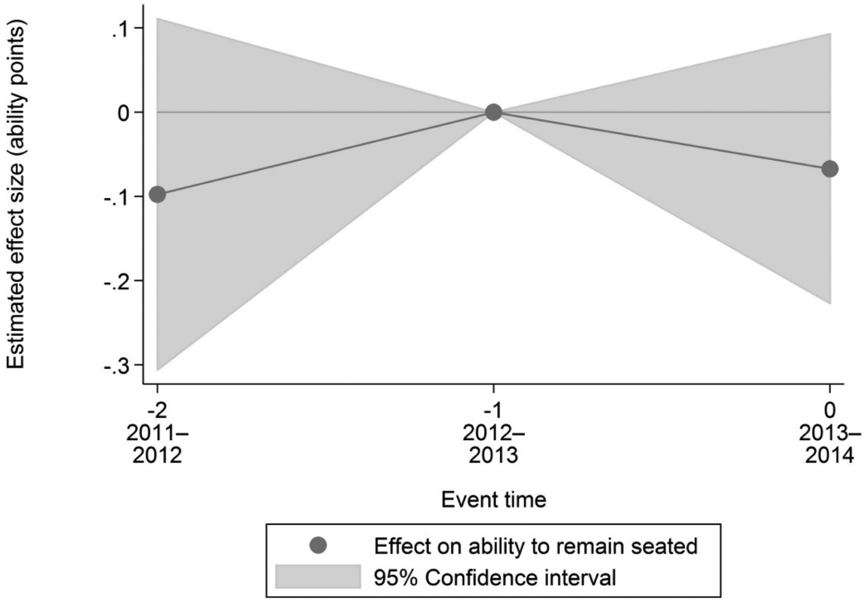
Notes: The figure shows the smoothed local polynomial graph of the results of ordinary least squares regression analyses across the pre-treatment school performance distribution in 2012/2013. Pre-treatment school performance is a moving window of 200 children. The dependent variable in each regression is time spent on physical activity during school-time. The estimated effect size based on difference-in-differences estimations is plotted on the y-axis. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure D4, panel B).
Source: Merged school performance and physical activity data.

Figure D5. The Effect of the Active Living Program on Time Spent on Physical Activity During School Time, by Pre-Treatment School Performance.



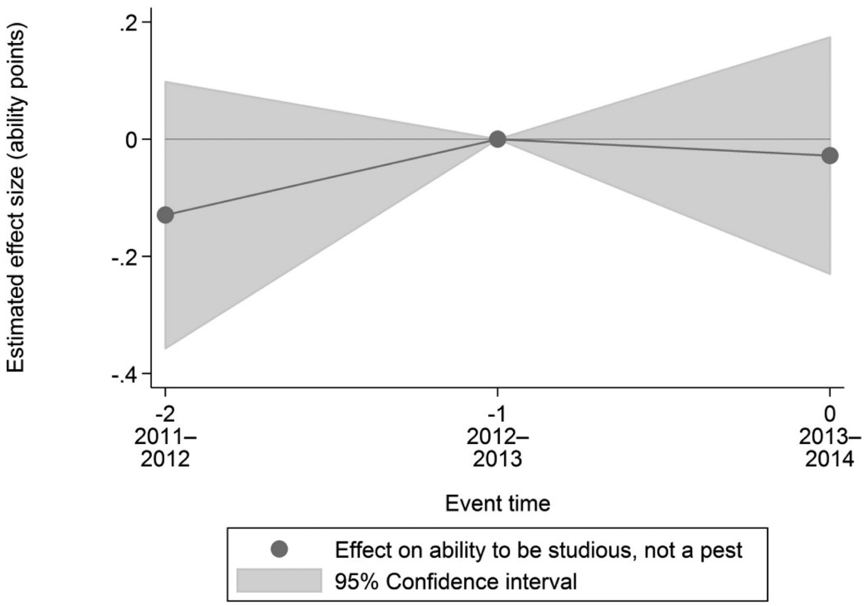
Notes: The figure shows the smoothed local polynomial graph of the results of ordinary least squares regression analyses across the pre-treatment overall school performance distribution. Pre-treatment school performance is a moving window of 200 children. The dependent variable in each regression is time spent on physical activity during leisure time. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables, and child fixed effects. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure D4, panel B).
Source: Merged school performance and physical activity data.

Figure D6. The Effect of the Active Living Program on Time Spent on Physical Activity During Leisure Time, by Pre-Treatment School Performance.



Notes: The figure shows the results of an ordinary least squares regression analysis. The dependent variable is children's self-assessed ability to stay in their seat when the teacher wants them to (on a scale from 5 to 8). The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are cohort fixed effects and a treatment dummy. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: Onderwijs Monitor Limburg questionnaire.

Figure D7. The Effect of the Active Living Program on Children's Ability To Stay in Their Seat When the Teacher Wants Them To.



Notes: The figure shows the results of an ordinary least squares regression analysis. The dependent variable is children's ability to be studious and not a pest (on a scale from 5 to 8). The estimated effect sizes based on a difference-in-differences estimation are plotted on the y-axis. Independent variables are cohort fixed effects and a treatment dummy. Adjusted 95 percent confidence intervals for estimated coefficients with robust standard errors are calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors.
Source: Onderwijs Monitor Limburg questionnaire.

Figure D8. The Effect of the Active Living Program on Children's Ability To Be Studious and Not a Pest.

Table D1. Interventions in the treatment schools.

	Treatment school 1	
Active transportation to school	• School pedestrian crossing indicators	• Put up signs at busy or unsafe roads in the school environments to indicate to children where they can cross the road safely on foot. These signs also signal to car drivers that they should pay extra attention because children may cross the road
	• Create traffic circle in schoolyard environment	• Area in the schoolyard where children can learn how to safely participate in different traffic environments, including one-way streets, roundabouts, zebra crossings, and traffic signs (see Picture D1, panel A)
	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
PA in school	• Establish ball game area	• Markings as indicator of a ball games area and regulations where to play ball games in the schoolyard
	• Put a ball backstop beside railway	• Re-facilitate the use of balls at the schoolyard by placing tall fences (see Picture D1, panel C) that prevent balls from falling on the railroad tracks next to the schoolyard
	• Additional sports day in schoolyard	• Introduce one additional sports day per school year
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
	• Prize contest for best idea to encourage PA	• Children asked to submit ideas about what they think could encourage them and/or their classmates to become more physically active
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports activities provided directly after school, e.g., gymnastics, soccer
	Treatment school 2	
Active transportation to school	• Availability of bicycle racks	• Provide bicycle racks
	• School pedestrian crossing indicators	• Put up signs at busy or unsafe roads in the school environments to indicate to children where they can cross the road safely on foot. These signs also signal to car drivers that they should pay extra attention because children may cross the road

Table D1. Continued.

	• Create traffic circle in schoolyard environment	• Area in the schoolyard where children can learn how to safely navigate different traffic patterns, including one-way streets, roundabouts, zebra crossings, and traffic signs. (See Picture D1, panel A)
	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
PA in school	• New fixed equipment in schoolyard	• Place new soccer goals
	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Playground markings	• Paintings, hopscotch markings, twister field
	• Establish ball game area	• Markings as indicator of a ball games area on the schoolyard
	• Sound equipment in schoolyard environment	• Install music equipment in the schoolyard to facilitate dancing
	• Additional sports day in schoolyard	• Introduce one additional sports day per school year
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
	• Use of schoolyard games	• Media cards that provide information to children on how to play games
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer
	Treatment school 3	
Active transportation to school	• Development safe route to school	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides a guided crossing where the crossing of a dangerous road could not be avoided
	• Mobilize crossing guards	• Approximately 15 minutes before and after school hours, crossing guards help children cross the road at the nearest (busy) road in the school environment
	• Adapt unsafe intersection in school environment	• Establish a new priority situation at an unsafe intersection close to the school
	• Create safer parking situation around school	• Redesign parking lots into parallel parking lots
	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike

Table D1. Continued.

	• School stimulation documentation on safe active transport	• Develop an active school transportation policy
	• Speed check action performed by children	• Children check the speed of car users in their school environment with the help of the local police. They also provide feedback to the drivers about their driving behavior
PA in school	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Additional sports day in schoolyard	• Introduce two additional sports days per school year
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
	• Use of schoolyard games	• Media cards that provide information to children on how to play games
PA in leisure-time	• Establish training circuit	• A "green" track next to the schoolyard (see Picture D1, panel B)
	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer
	• Establish school soccer team	• Establishment of a school team that regularly plays soccer after school, available for all interested children, irrespective of their gender
	• Establish PA activities by children for local residents	• Establishment of "Neighborhood in Action" group that encourages physical activity by performing family (physical) activities, such as walks through parks
	Treatment school 4	
Active transportation to school	• Develop safe route to school	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides guided crossings where the crossing of a dangerous road could not be avoided
	• Mobilize crossing guards	• Approximately 15 minutes before and after school hours, crossing guards help children cross the road at the nearest (busy) road in the school environment
	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
	• School stimulation documentation on safe active transport	• Develop an active school transportation policy
	• Speed check action performed by children	• Children check the speed of car users in their school environment with the help of the local police. They also provide feedback to the drivers about their driving behavior

Table D1. Continued.

PA in school	• New fixed equipment in schoolyard	• Place a climbing structure in the schoolyard
	• Additional sports day in schoolyard	• Introduce two additional sports days per school year
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer
	Treatment school 5	
Active transportation to school	• Develop safe route to school	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides guided crossings where the crossing of a dangerous road could not be avoided
	• Lessons to improve bicycle skills	• Several skill lessons on how to safely ride a bike
PA in school	• New fixed equipment in schoolyard	• Place new soccer goals in the schoolyard
	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer
	Treatment school 6	
Active transportation to school	• Availability of bicycle racks	• Provide bicycle racks in front of the school
	• School pedestrian crossing indicators	• Put up signs at busy or unsafe roads in the school environments to indicate to children where they can cross the road safely on foot. These signs also signal to car drivers that they should pay extra attention because children may cross the road
	• Introduce "Walk/Bike to school day"	• Mark one day in the week or month as active transportation day

Table D1. Continued.

PA in school	• New fixed equipment in schoolyard	• Place a table tennis court in the schoolyard
	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Playground markings	• Paint hopscotch markings on the schoolyard
	• Establish ball game area	• Introduce regulations regarding where to play ball games in the schoolyard
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	Treatment school 7	
Active transportation to school	• Sticker competition for active transport	• Provide stickers to children who traveled to school on foot or by bike
PA in school	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Playground markings	• Paintings, hopscotch markings
	• Use of schoolyard games	• Media cards that provide information to children on how to play games
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	Treatment school 8	
Active transportation to school	-	-
PA in school	• New fixed equipment in schoolyard	• Place soccer goals, a climbing structure, and a table tennis court in the schoolyard
	• Establish ball game area	• Introduce regulations regarding where to play ball games in the schoolyard
	• Additional sports day in schoolyard	• Introduce one additional sports day per school year
	• Use of schoolyard games	• Media cards that provide information to children on how to play games
	• Prize contest for best idea to encourage PA	• Children asked to submit ideas about what they think could stimulate them and/or their classmates to become more physically active
PA in leisure-time	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Table D1. Continued.

	Treatment school 9	
Active transportation to school	• Develop safe route to school	• A marked route between the school and the closest public playground that avoids dangerous crossings and provides guided crossings where the crossing of a dangerous road could not be avoided
	• Lessons to improve bicycle skills	• Several skill lessons on how to safely ride a bike
PA in school	• New fixed equipment in schoolyard	• Place a climbing structure in the schoolyard
	• New loose equipment in schoolyard	• Provide balls in different sizes and colors, hula hoops, skipping ropes, and other small, loose playground equipment
	• Sports clinics in recess	• Local sports clubs provide trial lessons and information during recess
PA in leisure-time	• Active Living Games	• An additional sports day for all participating treatment schools (funded by the Province of Limburg)
	• Establish out-of-school PA program	• Multiple sports disciplines provided directly after school, e.g., gymnastics, soccer

Table D2. Differences between children in treatment and control schools in the pre-treatment year (2012/2013).

	Treatment		Control			
	mean	sd	mean	sd	Diff.	P-value
Panel A. School performance data						
Number of tests taken	4.92	(0.69)	5.09	(1.12)	−0.17	0.003
Overall school performance	57.67	(7.73)	56.86	(8.06)	0.81	0.103
Language performance	57.31	(7.77)	56.63	(8.02)	0.68	0.169
Math performance	56.49	(7.72)	55.56	(8.27)	0.93	0.065
Age when tests were taken (months)	125.51	(8.05)	126.75	(8.63)	−1.24	0.018
Gender (1 = boy)	0.48	(0.50)	0.47	(0.50)	0.00	0.946
Cohort (1 = 5th graders, 0 = 4th graders)	0.55	(0.50)	0.60	(0.49)	−0.05	0.095
Panel B. PA data						
Number of days the accelerometer worn	3.78	(1.32)	4.07	(1.27)	−0.28	0.000
Time spent on PA during school time (min./day)	85.29	(25.14)	100.28	(25.44)	−14.99	0.000
Time spent on PA during leisure time (min./day)	201.92	(52.33)	200.51	(47.24)	1.41	0.712
Age during PA measurement (months)	128.14	(8.21)	128.47	(8.27)	−0.33	0.468
Gender (1 = boy)	0.49	(0.50)	0.41	(0.49)	0.08	0.002
Cohort (1 = 5th graders, 0 = 4th graders)	0.47	(0.50)	0.50	(0.50)	−0.03	0.266
Wearing time of accelerometer (min./day)	419.87	(137.63)	429.56	(128.12)	−9.69	0.176

Notes: The table shows results from *t*-tests on the equality of means.
Sources: School performance data and PA data.

Table D3. Difference-in-differences estimations of the effect of the Active Living Program on school performance on language and math tests.

	(1)	(2) Language	(3)	(4)	(5) Math	(6)
Dependent variable: school performance on language or math tests	OLS	FE	FE	OLS	FE	FE
Treatment * Post	−1.231	−1.004	−1.072	−1.191	−1.216	−1.309
	(0.595)	(0.575)	(0.586)	(0.568)	(0.549)	(0.511)
Adjusted p-value	[0.055]	[0.099]	[0.087]	[0.052]	[0.041]	[0.022]
Treatment (1 = Treatment group)	0.682			0.929		
	(0.821)			(0.859)		
Post (1 = 2013/2014, 0 = 2012/2013)	7.979	7.832	3.892	8.409	8.444	−0.448
	(0.479)	(0.459)	(3.844)	(0.480)	(0.450)	(3.442)
Age when tests were taken (months)			0.803			1.250
			(0.418)			(0.285)
Age squared			−0.002			−0.002
			(0.001)			(0.001)
Child fixed effects	No	Yes	Yes	No	Yes	Yes
Constant	56.630			55.561		
	(0.574)			(0.455)		
Observations	2,064	2,028	2,028	2,057	2,016	2,016
R-squared	0.175	0.751	0.754	0.200	0.783	0.789
Number of children in the estimation sample	1,050	1,014	1,014	1,049	1,008	1,008
Mean school performance in estimation sample	60.722	60.640	60.640	60.015	59.938	59.938
Standard deviation	(8.814)	(8.784)	(8.784)	(8.731)	(8.724)	(8.724)
Mean school performance in full sample	61.623	61.623	61.623	60.964	60.964	60.964
Standard deviation	(9.248)	(9.248)	(9.248)	(8.882)	(8.882)	(8.882)

Notes: The table shows the results of six ordinary least squares regression analyses. The dependent variable in columns 1 through 3 is school performance on language tests. The dependent variable in columns 4 through (6) is school performance on math tests. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Source: School performance data.

Table D4. Descriptive statistics of the mechanism variables across cohorts.

		Control schools			Treatment schools		
	Cohort	mean	(sd)	N	mean	(sd)	N
Good at: Being calm and quiet	2011–2012	7.02	(0.84)	256	6.89	(0.79)	255
	2012–2013	6.87	(0.88)	223	6.94	(0.87)	235
	2013–2014	7.08	(0.84)	288	6.93	(0.79)	285
	2014–2015	7.01	(0.91)	176	6.81	(0.83)	226
Good at: Staying in my seat	2011–2012	7.16	(0.78)	255	7.06	(0.80)	253
	2012–2013	7.18	(0.79)	222	7.20	(0.75)	233
	2013–2014	7.21	(0.80)	287	7.18	(0.83)	282
Good at: Being studious, not a pest	2011–2012	6.91	(0.78)	255	6.78	(0.77)	255
	2012–2013	6.83	(0.89)	221	6.86	(0.82)	232
	2013–2014	6.86	(0.83)	287	6.89	(0.79)	282
Music lessons (1 = yes)	2012–2013	0.09	(0.29)	214	0.16	(0.37)	225
	2013–2014	0.13	(0.33)	282	0.13	(0.34)	278
	2014–2015	0.14	(0.34)	175	0.13	(0.33)	224

Source: Onderwijs Monitor Limburg questionnaire.

Table D5. Regressions of overall school performance on mechanism variables.

Dependent variable: overall school performance	(1) Being calm & quiet	(2) Staying in seat	(3) Being studious & not a pest	(4) Music lessons
	1.535	1.612	2.155	2.847
	(0.189)	(0.187)	(0.196)	(0.404)
Adjusted p-value	[0.000]	[0.000]	[0.000]	[0.000]
Constant	57.857	57.037	53.821	68.238
	(1.339)	(1.363)	(1.355)	(0.282)
Observations	3,181	3,164	3,151	3,116
R-squared	0.027	0.027	0.047	0.016

Notes: For this table, only non-Active Living schools are selected. Results are from four ordinary least squares regression analyses. The dependent variable in each column is overall school performance in sixth grade in 2013/2014. The independent variables are (on a scale from 5 to 8): children's self-assessed ability to be calm and quiet when the teacher wants them to (column 1), children's self-assessed ability to stay in their seat when the teacher wants them to (column 2), and children's self-assessed ability to be studious and not a pest (column 3). The independent variable in column 4 is whether a child takes music lessons (1 = yes). Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted *p*-values for estimated coefficients with robust standard errors, calculated using a *t*-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.
Sources: School performance data and Onderwijs Monitor Limburg questionnaire.

Table D6. Years in which mechanism questions were asked.

		School year			
		2011/2012	2012/2013	2013/2014	2014/2015
Active Living cohorts	grade →	3	4	5	**6**
		4	5	**6**	
Pre-Active Living cohorts	grade →	5	**6**		
		6			

Legend:

- (white box) Grade in which children did not fill out a questionnaire
- **6** (thick-bordered box) Grade in which the pre-treatment cohorts filled out the OML questionnaire
- **6** (black box) Grade in which the treated cohort filled out the OML questionnaire, during the post-treatment year 2013/2014 (i.e., those who were enrolled in grade 5 in the pre-treatment year 2012/2013)
- **6** (grey box) Grade in which the treated cohort filled out the OML questionnaire, 1 year post-treatment (i.e., those who were enrolled in grade 4 in the pre-treatment year 2012/2013)