The effects of public high school subsidies on student test scores*

The case of a full-day high school program in Pernambuco, Brazil

Leonardo Rosa†‡ Eric Bettinger§ Martin Carnoy† Pedro Dantas¶

February 26, 2020

Working Paper

Please click here for the latest version

Abstract

Education in Brazil and Latin America is typically associated with limited school daily hours. This paper explores a full-day policy in Brazil to assess the impact of extending high school day on student test scores. The full-day program we evaluate increased the number of hours for math classes by 50% and by 20% for language classes and provided more resources to participants' schools. The implementation of the program generates variation within and between schools, which allows us to estimate a triple differences model and separated effects of more hours for students. Our estimates show that over three years, the full-day school program increased math and language student test scores by 0.22 and 0.19 standard deviations, respectively. These impacts were equivalent to an increase in the gains of students by 50% in math and by 35% in language, compared to the average achievement growth of high school students in regular public high schools. To mitigate potential selection bias concerns, we also provide alternative methods to compute estimate effects. Our preferred alternative–an instrumental variable approach–shows that our results are very robust. Finally, we explore heterogeneity in the effects of the program according to changes in full-day school characteristics. Our results indicate that the number of days of extended hours at full-day schools is highly associated with greater effects.

*The authors gratefully acknowledge funding from the Lemann Center for Educational Entrepreneurship and Innovation in Brazil and the Department of Education of Pernambuco (Brazil) for their kind cooperation in providing the data for this research and helping us understand the details of the full-day school reform.

†Corresponding author. E-mail: **leonardo.rosa@stanford.edu**

‡Stanford University

§Stanford University and NBER

¶Somos Professores

1 Introduction

A major change in education worldwide over the past generation is the rapid expansion of post-primary schooling (Banerjee et al., 2013). In most countries of the world, more than 50 percent of youth attend secondary education. A second major change is that, thanks to more countries participating in international testing, there is greatly increased awareness that the amount of learning taking place in post-primary schooling varies greatly among national educational systems, and that, furthermore, it is not easy to improve student performance at this level of education (OECD, 2016). These changes have led many countries to focus on reforms that could meet the challenge of improving student achievement in middle and secondary schools.

One popular reform, especially in Latin America, is to increase the amount of time students are exposed each week to core subject matter, such as mathematics, language arts, and science. This emphasis on increased exposure is logical since in most Latin American countries students attend school –even secondary school—only a few hours daily. Brazil is one such country, and can serve as an important example of both the challenge in improving post-primary education and the impact of this major effort in the region to improve quality through more time in school.

In 2003, the average high school in Brazil delivered only 4.5 hours of education per day, well below the average of seven or more hours in most OECD countries that have PISA test scores (Bruns, Evans, and Luque, 2011). Increasing the number of school hours is a challenge in Latin America, however, because it is expensive. Furthermore, there is no consensus that extending the school day is beneficial for student achievement (Alfaro, Evans, and Holland, 2015). In Brazil, specifically, the evidence available shows that increasing school day hours without more time dedicated to math or language has null effects on student test scores (Almeida et al., 2016).

This paper aims to provide new evidence to this debate by analyzing a full-day school program in Pernambuco, Brazil. The program was implemented in high schools only, and it combined longer school day hours with increased resources for high schools. In this program, the state government converted regular (4.5 school-day hours) public schools[1] into full-day high schools (8 school-day hours). The extended hours were in core curricu-

[1]As in many schools in developing countries, Brazil hosts double-shift schools which have one set of students and faculty in the morning and a second set of students and faculty in the afternoon.

lum components (language, mathematics, and sciences): hours of instruction increased by 40-80 percent, depending on the school and subject. To illustrate the dramatic change in the amount of hours, Figure 1 shows the difference between weekly hours of mathematics in regular and full-day schools in Pernambuco compared to other PISA participants. While regular schools in Pernambuco have the fewest number of hours of mathematics per week across all 58 PISA countries listed, the full-day schools are among the 20 top countries with the most hours of mathematics per week.

[Figure 1 about here.]

As one of their central educational policies, the state government promoted the expansion of the program, which was accompanied by increases in the state's performance in national test rankings. These results drew the attention of other state and federal authorities, which culminated in the federal government launching a national full-day school program.

In this paper, we address three fundamental unanswered questions about the full-day school program in Pernambuco. First, what are the key components of the full-day school program in Pernambuco? By answering this question, we try to understand the critical elements of the program, including a better understanding of the extended hours. Second, does the full-day school program affect student test scores? Although this question seems crucial from a policy standpoint, there is no previous study that tries to establish a causal link between the full-day program and students' test scores. Third, are different components of the program mediating the effects on student test scores? To answer this question, we explore variation in the program components, focusing mainly on the different amount of extended days each school implemented.

The paper begins by presenting a comprehensive description of the program. In Section 2, we use official documents and different data sources to systematize the critical components of the program. We argue that although there are several changes to school structure and human resources, the changes in time for core curriculum subjects is the crucial component of the program. Furthermore, in Section 2 we give special attention to the implementation process of the program since this process is fundamental for our empirical strategy to identify the effects of the program on student test scores.

In light of the full-day school program's implementation and the data we have, Section 4 turns to our empirical strategy for estimating the effects of the program on student math and language test scores. A threat to identifying the program's effects is the selection of students: students attending the full-day program might be different from students attending regular schools. We address this threat by combining new data with causal inference methods. With respect to the data, described in 3, we use administrative records to construct a new dataset that tracks individual students' test scores before and after the program's implementation. Therefore, all our results can be seen in terms of student gains over their high school years. By doing this, we aim to provide new evidence using panel data and contribute to previous literature, which has analyzed the effects of full-day programs in developing countries on student test scores using only cross-sectional data (Bellei, 2009; Almeida et al., 2016).

Controlling for pre-program test scores might not solve the selection problem: students attending full-day schools might also have larger test score gains than students attending regular schools even without the program. As a step toward mitigating the selection problem, our research design explored variation in the implementation of the full-day program. In short, the implementation of the full-day program creates two types of variation in treatment status: within and between schools. Within schools, only students attending high school in the year of the program's implementation get the full-day program; the cohorts attending the school before the conversion do not get the program. Thus, we can define the treatment group as those exposed to the program and the control group as those that are not exposed to the program within the same school. This variation in full-day school exposure allows us to implement a difference-in-differences approach comparing the results of treatment and control students before and after high school. Because the program was not implemented in all high schools, we can also explore variation in program exposure between schools. Thus, we compare the gains of students in the same cohort who remained in regular schools[2]. Combining within- and between-school variation, we then estimate the effects of the full-day school program using a triple differences approach. Our strategy is based on the idea that cohorts before the program in a particular year form a useful counterfactual for cohorts enrolling in the full-day program, after accounting for fixed differences between the schools and for common time effects.

Finally, if selection happened at the same time the program is implemented, the triple

[2] a "naive" or "placebo" DD, as described in Asim and Dee (2016)

differences method does not eliminate the bias. To check the robustness of the triple differences approach, we also implement an instrumental variable approach. Using data from before the program started, we constructed an instrument based on a student's probability of attending a regular or full-day high school conditional on the distance from the student's middle school. Because these probabilities were constructed with pre-program data, they are unlikely to be "contaminated" by selection of students. Once we combined the instrument variable with the difference-in-differences approach, we followed the procedure suggested by Bettinger et al. (2017). Thus, our identification strategy exploits two key influences on students' attending full-day school: (i) the change in the availability of the full-day program from cohort to cohort, and (ii) the distance between each middle school and available high schools. Altogether, our exclusion restriction requires that (i) any other mechanism through which student's distance from middle to high school affects test scores is constant across cohorts; and (ii) any other mechanism causing test scores to differ between cohorts with and without a full-day school exposure will affect students homogeneously after accounting for their probability of attending a particular high school.

Our results suggest that attending a school converted into a full-day high school leads to large and significant gains in student test scores. Relying on within-school variation in the timing of exposure to the program, we estimate that spending three years in the program leads to an average of 0.22 standard deviations increase in standardized math scores and a 0.19 in language (Portuguese) test scores. In addition, differentiating between types of full-day high schools, we show that two additional days of full-day school (3 vs. 5 full-days), increases test scores by 0.18 sd in math and by 0.10 sd in language. We argue that these differences are plausibly connected to the use of time for academic activities since the infrastructure and human resources in these two types of full-day schools are very similar. Finally, our results are robust to several checks, including the instrumental variable approach described above.

Our research contributes to two strands of literature. First, it provides new evidence on the effects of extra time on academic activities on student test scores. Prior research has used different approaches to study these effects. For example, some macro-level evidence suggests that time on a given task matters when explaining differences between countries test scores, as suggested by Lavy (2015) and Rivkin and Schiman (2015). Studies more similar to ours, however, are those that analyze the effects of particular interventions focused on the relationship between the time on academic activities and student test scores (Huebener, Kuger, and Marcus, 2017; Battistin and Meroni, 2016; Taylor, 2014; Cortes and Goodman, 2014; Agüero and Beleche, 2013; Meyer and Van Klaveren, 2013; Bellei, 2009). Most studies have focused on interventions in developed countries, with a few exceptions, notably Bellei (2009), Agüero and Beleche (2013), and Cabrera-Hernández (2020). We therefore also aim to contribute to understanding the importance of school day hours as a potential input for improving education in developing countries.

Second, this paper adds to the literature on the complementary effects of expanding school resources and task activities on student test scores. Lavy (2016), for instance, has similar conclusions to our study. Lavy (2016) finds positive results of an educational intervention in Israel that combines resources with additional time on academic activities. Our study also indicates, in a very different (and financially constrained) context, that interventions like the full-day school program are successful in increasing students test scores.

2 The full-day school policy in Pernambuco, Brazil

2.1 High school in Pernambuco

The state of Pernambuco is located in the Northeast of Brazil, one of the poorest regions in the country. Pernambuco's per-capita GDP is nearly \$6,000 USD, which is smaller than the average per-capita GDP in Brazil (close to \$9,800 USD). The state's population is nearly 9 million, and over 35% are between 4 and 19 years old, the school age in Brazil.

This study focuses on students attending high schools in Pernambuco (age 15-19). Between 2005 and 2017, the state's average score on the national high school student achievement indicator (*IDEB*) increased from 2.7 to 4.0, on a scale from 0 to 10. While the average score is still low, the state that used to have one of the ten lowest scores in 2005 across all states became one of the top three highest scoring states in Brazil in 2017. An explanation for this progress, according to state authorities, the press, and others, is the implementation of the full-day high school program. Because of Pernambuco's high school achievement gains, the program attracted broad interest and even inspired a full-day program for high schools supported by the federal government.

Although the full-day program has garnered much attention, few studies have analyzed the program, and there is no evidence that the program has a causal effect on student test scores. This study aims to fill this gap by estimating the effects of Pernambuco's full-day high school program on student test scores using a causal inference approach.

2.2 Full-day high school program: implementation

We believe that the implementation of the program is unique and helpful to understand our research design choices. Furthermore, the program has components that go beyond an increase in school day hours only. In this section, we describe the implementation of the program and provide details about additional components that are part of the program, including details about the use of the extra school day hours.

In 2004, Pernambuco's government started a pilot program to convert regular schools into full-day high schools. The government focused only on high schools, and they reorganized schools and provided resources to lengthen the the number of school day hours.

In the implementation process, we observe variation in the exposure of students to the program over time, geographical areas, and even across cohorts within schools. With respect to variation over time and geographical area, the government smoothly scaled up the program across years and municipalities. The goal of the government authorities was to achieve at least one full-day school per municipality; according to data shown in Figure 2, they achieved this goal over ten years.

[Figure 2 about here.]

The government did not convert all schools into full-day high schools. They aimed to maintain (at least some) regular high schools for students who did not want or could not attend a full-day high school. We checked this claim by comparing the enrollment in full-day and regular high schools. Figure 3b reports that, in 2014, about 40% of students starting high school attended a full-day school building. This result confirms that some students attended regular high schools while others attended full-day high schools. In practical terms, these two types of schools continuing to exist together allow us to estimate a model comparing students in full-day and regular high schools, after accounting for common year and school fixed effects.

[Figure 3 about here.]

The implementation of the program also created variation in exposure to the program across school cohorts within a school. We can describe a typical implementation schedule in one school as follows. In year $(t-1)$, the state announced the school would become a full-day school in year t. In year t, only those students enrolling in 10th grade (the first grade of high school in Brazil) attend the full-day program, i.e., get the extra school day hours. Students who attended 10th grade in year $(t-1)$ and enrolled in 11th grade in year t would continue in the same building, but they did not get the extra hours of school. They did, however, get other potential "benefits" of the program, such as the reforms in the infrastructure of schools and changes in human resources. The way schools implemented the full-day program generated two natural groups within the same school: a treatment group that received the full-day high school program entirely, and a control group of students who attended the same building but did not receive the extra hours (or other features of the program that might be exclusive for the new cohort).

To check whether the implementation schedule within schools happened as described, we analyzed school day hours data reported by schools to the Ministry of Education (INEP, 2019b). Figure 4a compares average school day hours of regular schools and full-day schools before and after the program implementation focusing on students attending 10th grade (the first year of high school). The results show a sharp change in the school day hours at the year of implementation of the full-day school. The difference in school day hours between regular and full-day schools continues years after the full-day program's implementation, and regular schools do not seem to increase their school day.

[Figure 4 about here.]

To verify whether succeeding cohorts in full-day high school buildings had a different exposure to the program in terms of school day hours, we analyze the average school day hours, and compare different cohorts (grades) within the same schools. Figure 4b shows the average school day hours for all high school grades (10th, 11th, and 12th) before and after the full-day school program, only for schools that received the program. The results confirm that in the year of implementation, the school day hours increased only for high school first-grade (10th grade) classrooms. In the year after the implementation, the school day hours increased for high school first- and second-grade (10th and 11th grades) classrooms. Finally, three years after the program started, the school day hours increased for all high school grades. In sum, these results confirm that the implementation of the program followed the state government rules.

2.3 Full-day school program: changes in inputs

Figure 4 in the previous section shows the school day hours increased in the schools converted into full-day high schools. In this section, we aim to qualify the changes in school day hours and verify other components of the program.

2.3.1 Changes in time and curriculum

There are two types of full-day schools: some offer full-day activities three days per week, and other schools offer full-day school activities five-days per week. Overall, we refer to schools with either program as full-day schools and differentiate when necessary. While the increase in hours was similar in both types (3 and 5 days), there are differences in the curriculum that are important to highlight.

Table 1 compares regular and full-day public high school timetables: the number of hours in full-day schools increased by 30% in language classes and 50% for math and science classes, compared to regular public schools. The changes in the curriculum were accompanied by the introduction of new courses that were not taught in regular high schools. They include a "life-project" program where students worked on topics such as community values, financial education, and how to run a business. For full-day schools that were 5 days per week, there was also time reserved for supportive activities such as independent studies and remedial education. Altogether, we see that the changes in curriculum were most used to support academic activities, and because of that, we expect positive results on students' test scores.

[Table 1 about here.]

2.3.2 Changes in human resources

Principals

Between 2011 and 2014, Pernambuco's department of education replaced principals in many regular and full-day schools. To examine these changes, we analyzed a survey applied in most schools (INEP, 2019a). Table 2 shows the average (unconditional) characteristics of principals in regular and full-day schools. Results indicate that full-day school principals have fewer years of experience as principals and fewer years of experience in their current schools. These two results might indicate that principals in the full-day schools are more likely to be replaced than principals in regular schools. Table 2 compares other principals' characteristics as well. Results indicate that principals in full-day and regular schools are very similar in demographic characteristics. Full-day school principals, however, are more likely to be recruited through a competitive hiring process, and their wages are also higher than the wages of principals in regular schools.

[Table 2 about here.]

Teachers

The government did not require full-day schools to change their teachers, but teachers needed to agree to work exclusively for full-day schools given the extra school day hours. The government, moreover, paid higher wages for teachers to stay in the school (*gratificação*). Once in full-day schools, the teacher would teach any classroom, i.e., cohorts attending full-day program and cohorts not attending it. We do not have any information that suggests that better teachers were allocated to new full-day classrooms.

Table 3 reports teacher characteristics in regular, full-day high schools before the program and full-day schools after the program. Teachers in full-day schools after the program started are similar to teachers in regular schools and teachers in full-day school before the program started. However, there are some differences: notably, for example, teachers in full-day schools after the program started are more likely to teach fewer classrooms.

[Table 3 about here.]

2.3.3 Changes in grade configuration

Many regular schools are multiple level schools. Thus, they enroll students in middle and high school grades. The full-day high school program changed this configuration. Once the school was in the program, they should not admit new middle school cohorts. Over time, full-day schools would become exclusive for high school students. Using enrollment data by school and grade, we checked the magnitude of these changes. Results in Figure 5 suggest that while the enrollment in middle grades dropped considerably after the full-day school program was implemented, some full-day high schools still enrolled middle-grade students.

[Figure 5 about here.]

2.3.4 Changes in the school's physical inputs

The last relevant change in full-day schools was the reform (or construction) of spaces to be used for students. The government reformed labs and provided new inputs for schools. To investigate the size of these changes, we again used data from national surveys. These data provide information about the quantity and quality (scored by an enumerator) of school inputs. Results in Table 4 suggest that the infrastructure of the full-day schools was very similar to regular schools before the program started. Results also indicate some changes in school structure: for example, art labs and rooms for group study are more frequent after the program. Altogether, however, the data we analyze indicate only marginal changes in infrastructure.

[Table 4 about here.]

As we described, the full-day high school program might be seen as a package, and many changes in schools happened once the program started. In the next sections, we try to investigate the effects of this package. To do so, we present the data, our identification strategy to estimate the effects of the program, and the results. In the results section, we try to explore whether changes in some of the characteristics we showed in this section are associated with smaller or larger effects of the program on student test scores.

3 Data and Summary Statistics

3.1 Data sources

This paper combines three different administrative datasets to study the effects of the full-day high school program on students' test scores. From administrative enrollment records, which span the school years 2009-2016, we created a longitudinal dataset that tracks students' enrollment trajectories during their high school years. From these records, we obtained the following student information: a unique numeric identifier, name, age, gender, grade, and school.

We measure student test scores in math and language using administrative data from SAEPE[3], a state test applied by Pernambuco's Department of Education. Information from SAEPE span the school years 2009-2016 and include individual test scores for students in grades 5, 9, and 12. During the SAEPE exams, the Department of Education applies a contextual survey, which we use to construct student demographic and socioeconomic characteristics[4]. From SAEPE data, we recovered students' names and their school identifier, which we used to match with student enrollment data. We provided more details about the string matching process in Appendix B

Finally, from internal documents from the Department of Education, we identify the year that a school was converted into a full-day high program (2004-2014), and the type of full-day program the government implemented in the school–3 or 5 days of extended school day hours. After combining these data with the enrollment records, we assigned each student a treatment status, based on his or her cohort and school.

3.2 Sample restrictions

For the school years we analyzed, 692 thousand students attended 10th grade, i.e., the first year of high school. Our analytical sample, however, uses only 103,812 of those students. We imposed several restrictions (exclusions) that explain this difference: (i) schools that had fewer than 10 students enrolled in 10th grade or did not take part in the state test in

[3] *Sistema de Avaliação Estadual do Estado de Pernambuco*

[4] When we were not able to match a student's survey data to their survey responses, we defined all student/household characteristics as zero, and we always included a dummy for a missing characteristic in regressions that include these characteristics.

one of the years, (ii) schools that started the full-day program before 2011, (iii) cohorts in full-day schools if the cohort started before $(t-1)$ or after (t), and (iv) students in 10th grade whose test scores we could not observe before and after high school, including students who dropped out. This last restriction alone caused a considerable reduction in our sample, since dropout rates are close to 40-50% in the state. We include detailed information about the construction of our analytical sample in Appendix A. Moreover, in our analysis, we show evidence that student attrition does not affect our main results.

3.3 Summary Statistics

3.3.1 Analytical Sample

Table 5 shows descriptive statistics for the 103,812 students in our analytical sample. The average test score at the end of middle school (9th grade) in our sample is about 0.16 standard deviations higher than the average.[5] Students are more likely to be female, and minorities are over-represented. About 25% of students had a mother who completed secondary schooling, and almost 75% of the sample reported having a member in the household on welfare benefits from Brazil's conditional cash transfer program (*bolsa família*).

[Table 5 about here.]

3.3.2 Students in full-day and regular schools

Table 6 breaks down the summary statistics for students by school type. Columns 1-3 report the averages and standard deviations (square brackets) for students attending regular schools, full-day schools before the program started, and full-day schools after the program started. Column 4 shows the *p-value* of a comparison between the treatment group–students attending full-day high schools after the program started–and other students, after accounting for year and municipality fixed effects.

Results reported in Table 6 suggest that the characteristics of students in full-day schools after the program started are very similar to those of students attending regular

[5] We standardized test scores to have an average of zero and a standard deviation of one using all students in the ninth grade

schools. Full-day school students in the treatment group, however, are younger and have higher initial test scores (9th grade) than other students. As we discuss in the next section, our research design tries to account for these differences and, thus, mitigate threats to identification mostly related to selection bias concerns.

[Table 6 about here.]

4 Identification Strategy

4.1 Triple differences

To identify the causal effect of the full-day program on student test scores, we leverage variation in the timing of the full-day program's implementation and implement a triple differences model. Table 7 illustrates our model. As shown in the Table, we begin by comparing student test score outcomes before and after high school for students treated in full-day schools. Because many other factors can explain students' gains over their high school years, we use as a first control group students attending a full-day school building but who did not get the treatment—students who enrolled in the high school before the full-day program started. We believe they form a useful counterfactual for treatment students in the full-day high schools, because we can use them as a counterfactual to account fixed differences over time and mitigate potential selection happening in full-day schools. Altogether, analyzing the gains over the high school period and the comparison between the treatment and control groups in full-day schools, allows us implement a standard difference-in-differences model.

[Table 7 about here.]

We turn to our triple differences model by comparing the difference-in-differences (DD) in the full-day schools with one placebo DD model in the regular schools. The regular schools form a useful counterfactual because they did not receive the full-day program, but they did receive all other contemporaneous educational policies implemented in the state. Therefore, using this counterfactual allows us to control for fixed differences in student gains across cohorts.

The triple-difference estimate of exposure to the full-day school program is based on the following specification:

$$\begin{aligned}
y_{igjt} = \beta_0 + \beta_1 \text{Grade}_{g=12} + \beta_2 \text{Cohort}_t + \beta_3 (\text{Cohort}_\text{t} \times \text{Grade}_{g=12}) + \\
\beta_4 \text{Full-day}_j + \beta_5 (\text{Full-day}_j \times \text{Grade}_{g=12}) + \beta_6 (\text{Full-Day}_j \times \text{Cohort}_t) + \\
+ \beta_7 (\text{Full-day}_j \times \text{Grade}_{g=12} \times \text{Cohort}_{t=t^*}) + X_i' \alpha + \varepsilon_{igjt} \qquad (1)
\end{aligned}$$

where, y_{igjt} is the test score (math or language) of student i, in grade $g \in \{9, 12\}$, in school j, in the cohort starting high school (10th grade) in year t. Our coefficient of interest is β_7, which measures the effect of the full-day school program over three years of high school on students' test scores. Notice that the parameter is the average treatment effect for students attending full-day schools, that take the state exam in grade 12, and were in the cohort that received the treatment ($t = t^*$).

This model focuses on three main variables. $\text{Grade}_{g=12}$ is a binary variable that equals one for student test scores at the end of high school (12th grade), and zero for test scores before high school (9th grade). Cohort_t is a fixed effect of cohort or year. Full-day_j is a dummy variable indicating if the high school the student attended in the first high school year is a school that receives the full-day program. Each coefficient associated with these variables captures a different aspect of students' test scores. Coefficient β_1 estimates the average gain during high school for students in regular schools. The parameter β_2 captures differences over cohorts in student quality (in terms of their test score) before starting high school. β_3 indicates the difference in test score gains through high school for different cohorts attending regular schools. The coefficients β_4, β_5, and β_6 capture the difference in the fixed difference in test scores of students starting high school in full-day high schools buildings before the program started.

Finally, as auxiliary controls, we included the following student socioeconomic and demographic characteristics: age, gender, race, an indicator that a mother completed secondary education, an indicator that a family is a recipient of a conditional cash transfer program (CCT), an indicator that the house has fewer than two restrooms, an indicator that the household owns a computer, and an indicator that the household owns fewer than 20 books at home. When presenting our estimates, however, we show results both with and without these covariates.

4.2 Instrumental Variable

In our triple difference approach, we identify the effects of the full-day school program by separating the differences in test score gains through high school between cohorts treated by the program and cohorts not treated. We aimed to mitigate potential biases when accounting for time, cohort, and school fixed effects—as described above. Selection bias concerns, however, will still be a problem if students who decided to attend the full-day high schools in the year the program started are those who would have larger test score gains in the absence of the program.

In response to these selection bias concerns, we propose an instrumental variable strategy similar to the approach used in Bettinger et al. (2017) and Bettinger and Long (2009). In this approach, we instrument the full-day school treatment—($\text{Full-day}_j \times \text{Grade}_{g=12} \times \text{Cohort}_{t=t^*}$), in equation (1)—with:

$$\begin{aligned} Z_{it} \quad = \quad & P[\text{Full-day}_{it}] = \\ & \sum_{j \in J} \big(P[\text{Full-day}_i | \text{Attended HS } j, Year = t] \times \\ & \qquad P[\text{Attended HS } j | \text{Attended any HS} \in J, year = 2008] \big). \end{aligned} \tag{2}$$

Here, we have the interaction of two probabilities. $P[\text{Full-day}_i|.]$ equals 1 if student i's high school j, was converted into a full-day school in the year t. The second probability is the probability of attending a high school j, which we constructed using a grouped conditional logit model (Guimaraes and Lindrooth, 2007) given by:

$$P[\text{Attended HS } j | \text{Attended any HS} \in J, year = 2008]\big) = \frac{\exp^{\alpha H_{gj}^{08}}}{\sum_j \exp^{\alpha H_{gj}^{08}}} \tag{3}$$

where,

$$\alpha H_{gj}^{08} = \alpha_0 + \alpha_1 dist_{gj}^{08} + \alpha_2 (dist_{gj}^{08})^2 + \alpha_3 (dist_{gj}^{08})^3 + \alpha_4 \text{Private}_j^{08} + \alpha_5 \text{Size}_j^{08} + \upsilon_{ij}. \tag{4}$$

The subscript g refers to the "group" a student is included in.[6] Here, the group g is defined by the middle school the student attended. Table D.1 in Appendix D illustrates the layout of our dataset. We have an identifier for a group the student belongs in, the high-school they potentially could attend, the frequency that counts the number of students in the group that went to the high school, and the characteristics of the high schools. In our logit model, we included: (i) $dist$ which is the distance from middle g to high school j (a third degree polynomial), (ii) $Private$, which indicates if the high school is private or not, (iii) $Size$ which indicates the size of the high school.

Importantly, we only use data from the year 2008 to compute these probabilities. That year has the advantage of being a year not affected by the treatment in the years we are using in our analysis. Remember that our analytical sample is restricted by cohorts attending high school between 2010 and 2014 and schools converted into full-day high schools across these years. The restriction was entirely based on test score data limitations; scores were only available starting in 2009. Our data relative to enrollment, however, starts in 2007. Therefore, we can observe transitions between middle and high school grades for schools when they were not affected by the program. These probabilities should not be related to full-day school adoption, which is useful for our instrument and to mitigate selection bias concerns.

4.2.1 Alternative Instrument: enrollment in a middle school converted into full-day high school

As an alternative instrument, we explore the fact that many schools offer middle and high school grade levels. Because of this, students attending a middle school in a building that was selected for being converted will be more likely to stay in the same school regarding their "ability". In this case, we can construct a "Wald" instrument that is specified as a dummy variable that equals one if the student attended middle school in the same building that was converted into a full-day high school, and zero otherwise.

[6] As shown by Guimaraes and Lindrooth (2007), this is equivalent to an individual fixed effect conditional logit model. The advantage for us is computational since we can transform the individual level data–which would have billions of possible combinations—into a manageable dataset and estimate these probabilities.

5 Results

Based on the non-parametric comparison we show in Table 7, Figure 6 reports the average (unconditional) test score gains over high school years of students attending regular and full-day high schools for cohorts before and after the program started. The results presented in Figure 6 indicate that one year before the full-day high school program started, the math test score gains during high school for students attending full-day high schools building and not treated were similar to math test score gains attending the regular schools. Nonetheless, after the full-day high school program started, students attending full-day high schools who participated in the program had larger gains than students attending regular schools. In this section, we scrutinize these results and aim to establish a causal relationship between the full-day high school program and students' mathematics and language test scores.

[Figure 6 about here.]

5.1 Results using the triple differences model

Table 8 reports the effect of the full-day high school program on student math and language test scores. Column 1 presents results for the baseline model as described in Equation 1: the results indicate that the full-day high school program increased math test scores by 0.22 standard deviations (sd). As shown in Columns 2 and 3 in Table 8, point estimates are very stable when we include municipality fixed effects and students' characteristics as additional controls. Column 4 aims to implement a more robust model, substituting the *Full-day* fixed effects of the triple differences model describe in Equation 1 by school fixed effects and their interactions with cohort and grade fixed effects. The estimates do not change when using this alternative specification. Finally, Column 5 presents estimates for the effects of the full-day program on students' language test scores (repeating the last specification showed in Column 4). The results suggest that the full-day high school program also had a positive impact on student language test scores of similar magnitude–0.2 standard deviation.

[Table 8 about here.]

Table 8 also shows other results that are illustrative for understanding the magnitude of the estimated effects. Students in regular schools, for example, gained 0.43 standard deviation in their math test scores, on average, over the course of high school years ($Grade_{g=12}$). Thus, a comparison of this "typical" gain with the effects of the full-day high school program suggests that the full-day high school program increases students' test score gains by 50% over their high school years. Interestingly, as shown in Table 1, the number of hours in math classes increased by 50% as well.

Results presented in Column 3—after accounting for student characteristics—in Table 8 reveals other interesting patterns. Students who attended full-day school before the program started do not have larger middle school test scores or higher test scores gains than students who attended regular schools. Furthermore, the middle school test scores of students who attended full-day school after the program started are higher than middle school test scores of students who attended regular school, but this difference is not statistically significant. If we believe that large differences in middle school test scores could raise selection bias concerns, then the results in Column 3 indicate that this concern is small.

Before we turn to the instrumental variables estimates, we estimate the effects of the program by year of implementation. We believe that if the selection of students is occurring, the effects of the program on students test scores should be lower in full-day schools that implemented the program later on. Our rationale is that as soon as the government carries out the program in more and more schools, the potential for selection decreases since the pool of "high growth achievement" students also decreases.

The effects per year of the full-day school program are shown in Table 9. Panel A in Table 9 reports the results for math, and it suggests that the effects of the full-day high school program on student math test scores are uniform over the different years of full-day school program implementation. When comparing the results in a joint test for equality, we cannot reject the null hypothesis that they are equal. Panel B in Table 9, however, shows a different pattern when analyzing the effects of the full-day high school program on student language test scores. The results indicate that the effects were larger for students attending schools that implemented the program earlier. The joint test for equality rejects the null hypothesis that all the effects are equal. Altogether, these mixed findings still raise concerns regarding selection bias.

[Table 9 about here.]

5.2 Instrumental variables results

We now estimate the effects of the full-day high school program using the instrumental variables approach described in Section 4. In short, results reported in Table 10 show that: (i) our instruments are highly correlated with the decision of attending a full-day school or not, and (ii) if we believe these instruments are credible ways to mitigate selection bias concerns, the effects of the full-day school program on student high school test scores are around 0.22 standard deviations for math and 0.18 for language. Therefore, the results are not very different when compared to the results reported using only the triple differences model.

[Table 10 about here.]

Table 10 analyzes the effects of the full-day program using each proposed instrument separately.[7] Columns 2 and 3 in Table 10 present the estimates using as an instrument the probability of attending a full-day school based on student's school of choices before the program started. As we explained in 4, we observe all school choices between 2008 and 2009, i.e., before the government implemented the full-day program in the schools that are part of our sample. Therefore, these probabilities should not be affected by the full-day program.

These probabilities are illustrative for various reasons. First, if we believe that the program changed students' school choices dramatically, and this is the root of the selection bias concerns, these probabilities should not be related to student participation in the program. Second, in case that the selection is not coming from changes in the school choice process, these probabilities have a high correlation with student participation in the full-day program, and our results would be still biased. Third, an intermediary situation is the one we believe is happening. The selection concerns might be relevant for some students who changed their school choices. In this case, these probabilities would

[7]In all of these estimates, we considered as the endogenous regression model the specification described in 1. Tables D.2 and D.3 in Appendix D show the results using specifications that include municipality and school fixed effects, as estimated in Columns 4 and 5 in Table 7. Importantly, the results follow the same patterns in all specifications.

help us to mitigate these concerns since it would attribute as zero their chance to get the program. For students who did not change their behavior because of the program, the probabilities we constructed would be informative, highly correlated with taking the program, and the 2SLS estimator would give us a non-biased estimate for the effects of the full-day school program on student test scores.

Results reported in Column 2 in Table 10 show a strong correlation between student treatment and the probabilities we constructed with data before the program and assigned for students enrolling in high school. Column 3, moreover, reports the results when using these probabilities as instruments to estimate the effects of the full-day high school program. The effect of the full-day school on student test scores would be nearly 0.31 sd for math, and 0.17 sd for language. Importantly, these results do not change our conclusions drawn from the triple differences approach.

Columns 4 and 5 in Table 10 present the estimates computed by using as the instrument the fact that the middle school some students attended was converted into a full-day high school, in their transition from middle to high school. This instrument is credible if we believe that students did not migrate from one middle school to another in order to increase their chances of entering a full-day school. Also, it is a useful instrument if attending the buildings of a full-day school as a middle school student enhances the likelihood of being treated by the program and it is not correlated to their previous achievement.

Results in Columns 4 and 5 in Table 10 suggest that attending a middle school that was converted into a full-day high school is highly related to being treated. The effects of the full-day school program, using this instrument, are positive, but the effects are not statistically significant for conventional levels. Compared to the triple differences approach, the point estimates dropped for math and are similar for language test scores. Surprisingly, this is the opposite of the results we estimate for the effects of the full-day school program using the previous instrument (Column 3).

Finally, Columns 6 and 7 in Table 10 combine the two instruments to estimate the effects of the full-day high school program on student test scores. Results reported in Column 6 suggest that the instruments combined present a stronger correlation with program participation than using each instrument separately. Results reported in Column 7 indicate positive effects of the full-day school program on student test scores. Furthermore, the magnitudes of these effects are similar to the effect sizes estimated using the triple differences approach alone. The standard errors, however, are larger. In our case, with two instruments we have an overidentified equation, so we may test whether the excluded instruments are appropriately independent of the error to evaluate the validity of the instruments. The p-values for this test are presented in Column 7 in Table 10. Results do not allow us to reject the null hypothesis that both instruments are uncorrelated with the error term.

In sum, the results we found using the instrumental variable approach suggest that the full-day high school program had a positive effect on student test scores, in math and language. In general, the magnitude of these effects is similar to the results we found using a triple differences approach alone.

5.3 Other Robustness checks

5.3.1 Parallel trends assumption

Changes in students' composition

In a triple differences framework, we expect that the differences between the groups are stable over time and that the changes in exposure to the full-day school program are not associated with changes in the distribution of covariates. To examine this aspect of the validity of the triple differences method empirically, we estimate regressions that use the characteristics of students that we can observe before and after high school as the dependent variables. Under the null hypothesis that there are no compositional changes, we expect that the full-day high school program had zero effects on these covariates.

Table 11 reports the results for this analysis. Overall, the full-day school program does not affect student composition. The only exception is the number of computers: treated students are more likely to have more computers between the 9th and 12th grade than other students. The magnitude of this effect, however, is not large considering that 50% of students in the control group had computers. Altogether, we believe this exercise contributes to the hypothesis that our triple differences model captures the effects of the full-day high school program and that other factors are not influencing our findings.

[Table 11 about here.]

Pre-trends from 5th to 9th grade

Ideally, we would like to examine the possibility that previous outcomes anticipate future treatment exposures. In our context, that means to estimate the effects of the full-day school program on student test score growth between elementary and middle school (5th and 9th grades). We expect that future full-day school program will not be associated with these gains. Although we cannot perform this test for all students in our entire sample sample[8], we can test it for students who start high school in 2014. Using this cohort, we compare students in regular schools and full-day schools using a standard difference-in-differences model. While the nonexistence of pre-trends in this restricted sample might not be sufficient, it is the best we can do to show that students present very similar trajectories before the program started.

Figure 7 presents results for the pre-trend test. It uses unconditional averages of test scores of students attending elementary, middle, and high schools. The results indicate that gains between elementary and middle school follow the same trajectory. After middle school, however, students who attended a full-day high school program had larger gains than students attending regular school.

Table D.4 in Appendix D presents a formal test for this analysis, using a differences-in-differences regression. The conclusions are the same: students who started high school in 2014 and attended the full-day high school program did not have a different test score gain between elementary and middle school than students who attended regular high schools.

[Figure 7 about here.]

5.3.2 Attrition

Finally, we investigate whether student attrition could affect our results. As we discussed in Section 3, we restricted our data to have only individuals who have test scores before and after high school. Now, we relax this restriction and analyze how different attrition patterns could bound our main findings.

We start by analyzing whether attrition rates are different across the control and treatment groups. As shown in Table D.5 in Appendix D, attrition is on average higher

[8]This is because the state test began in 2009.

for students who are low-achievers, older adolescents, and males. This pattern, however, is not different when we compare attrition patterns in regular schools and full-day high schools. Except for student age, attrition rates are very similar for both school types.

Next, we test how attrition could affect our results depending on students' test scores. We follow a similar approach to the one used by Angrist, Bettinger, and Kremer (2006). In this approach, we censored the data at the k-th percentile of the 12th-grade data. Anyone with a missing test score or a score at or below the k-th percentile were assigned the k-th percentile. Then, we estimated the effects of the full-day high school program for these censored samples.

Panel A in Figure 8 reports the results for these winsorized regressions. We find positive effects so long as values imputed are below the 75th percentile. If students who are dropping out or being held-back would have scored above the 75th percentile, then the estimated impact would be statistically indistinguishable from zero. For all other cases, however, the effects would be positive.

How we interpret the results shown of Panel A in Figure 8 depends on the potential outcomes of students who did not take the test. Although these potential outcomes are not observable, we can have a better idea based on the profile of attrition by the initial test scores of students who attrited. Panel B in Figure 8 shows the attrition rates by student test scores before the program. The pattern in Panel B suggests that students who attrited tended to come from the bottom of the distribution. If their potential gains would not change their position in the test score distribution, then the most likely situation would be one where the effects of the full-day school program are similar to the one we presented as our main results in Table 7, i.e., an effect near 0.2 sd.

[Figure 8 about here.]

5.4 Exploring Mechanisms

To shed light on the mechanisms through which full-day high school program affects student outcomes, we examine heterogeneity in the effects according to different changes in the schools that implemented the full-day program. As we discussed in section 2, the full-day high school program is a package that includes changes in infrastructure, human resources, and time at schools. To explore how variation in this package affects our results, in this section, we divided schools into groups, based on their differential changes in principal replacement, average characteristics of schools, and extra days of the full-day program in the week; we then analyzed the effects using a triple differences model.

5.4.1 Principal replacement

To determine how principals' replacement in full-day high schools is associated with student outcomes, we employ the triple differences model in subgroups of the full-day schools. We divided full-day schools into four groups: (i) full-day schools that did not replace their principal, (ii) full-day schools that replaced their principal before the program started, (iii) full-day schools that replaced their principal in the year the program started, and (iv) full-day schools that replaced their principals in the second or third year after the program started.

Table 12 reports the heterogeneous effects on students' math test scores by principal replacement status in the full-day high schools. The results indicate larger effects of the full-day school program on student test scores when students had a lower exposure to new principals. For instance, while the estimated effects of the program on math test scores for schools that did not replace their principals were nearly 0.3 standard deviation, estimated effects in schools that replaced their principals before the program started were a lower 0.18 standard deviation. As we show in Table D.6 in Appendix D, this pattern is similar when we use student language test scores as the outcome. These results might indicate that principals that have more experience in the school are more likely to better implement the program.

[Table 12 about here.]

5.4.2 Teacher Average Characteristics

We turn our analysis to explore how changes in the composition of teachers in the full-day school influence the effects of the program. To determine how different changes in teacher composition in full-day high schools are associated with student outcomes, we estimate the effects of the program using the triple differences model in subgroups of the full-day schools. For each full-day school in our analytical sample, we compute the average characteristics of teachers. After that, for each average teacher characteristics, we divide schools into two groups: below and above the median.

We explored six teacher characteristics on full-day schools: (i) the percentage of teachers who are female, (ii) the percentage of teachers who graduated from a public college (proxy for elite college), (iii) the percentage of new teachers at the school during the academic years in our analysis, (iv) the percentage of teachers who taught more than one cohort (proxy for being an exclusive teacher for full-day treated cohorts), (v) the average number of classrooms that teachers taught, and (vi) the average number of schools that teachers worked.

Table 13 reports the heterogeneous effects of the full-day school program on student math test scores. The results suggest that full-day high schools that had a larger proportion of female teachers, and teachers with a larger average number of classrooms, are associated with larger effects of the program on student math test scores. Similar patterns are found when we analyzed language test scores, as shown in Table D.7. Overall, however, changes in teacher characteristics are not associated with heterogeneity in the effects of the program. This lack of differential effects seem to suggest that changes in teacher composition are not driving the results.

[Table 13 about here.]

5.4.3 Extra days of full-day school with academic support activities

One of the most remarkable differences across full-day high schools is the number of longer school days during the week. The government implemented one of two variations of the program: full-day high schools with 3 days of longer school day hours and full-day schools with 5 days of longer school day hours. The two types differ in terms of extra-academic activities. As we show in Table 1, students in 5-days full-day schools had independent studies or remedial class in these two extra days. Therefore, students taking these classes would use their time to study and improve their academic skills, focusing on the content from the courses taken, such as math and sciences.

Table 14 reports the heterogeneous effects of the full-day schools on student test scores by school type. The results suggest that the effects of the full-day school program on student test scores were larger for students attending schools that implemented the longer school hours 5 days per week (0.4 sd in math and 0.26 sd in language) than for students attending schools with longer school hours 3 days per week (0.20 sd in math and 0.18 sd in language). Results in Table 14 also indicate that effects are larger for math test scores than language test scores. Altogether, these results suggest that the program is more effective when standard curriculum-based classes are combined with supportive activities, such as independent study and remedial classes.

[Table 14 about here.]

6 Conclusions

In this paper, we estimate the effect of school resources and additional hours in the school day on high school students' test score performance. We take advantage of a policy reform in Pernambuco, Brazil that began in 2004, which changed the rules of funding public high schools. The system changed from funding schools based on a 4-hour curriculum per day, multiple shifts, and multiple grade-level schools to a system that funded full-day schools working on an 8-hour curriculum and in buildings exclusively for high school students. We analyze the effects of this reform on student test scores by extensively exploiting the variation created by the government when implementing the program. This variation allows us to create treatment and control groups and estimate the effects of the program by using a triple differences model where we can compare treatment and control groups in the same schools. Furthermore, because we can observe student test scores before and after high school, we can analyze these effects from the perspective of gains made over high school years.

The results based on high school students' test score gains indicate that the full-day high school program has a positive and significant effect on student performance in math and language—0.22 and 0.2 standard deviations, respectively. Importantly, over the high school years, the full-day school program increased the math test scores of students by 50% and the language test scores by 35% when compared to the math and language test score gains of students attending regular high schools. This magnitude is symmetric to the 50% increase in time of math instruction and the 20% increase in time of language instruction in full-day high schools.

To mitigate concerns regarding students' selection into the program, we provided several additional estimates. Our most important alternative identification strategy was based on instrumental variables regressions that exploit as instruments: (i) the pre-program probabilities of attending full-day schools, and (ii) if the middle school the student attended was converted into a full-day high school. We argue that these instruments might be seen as credible since they affect the likelihood of being treated, and they are not directed related to the potential gains made by students over their high school years. Although the point estimates of the effects of the program varied using the instrumental variable estimates, the direction and magnitude were always similar to those estimates found using the triple difference method. On average, the full-day school program affected students' test scores by nearly 0.2 sd. Besides the instrumental variables, we perform other tests to challenge our findings, and these suggest that the results are very robust to attrition, changes in student socio-demographic composition, and pre-trends in test score gains.

We also test for heterogeneity in the effects of the program across different characteristics of full-day schools. We based our heterogeneity analysis on components of the program. Thus, we focused on aspects of full-day schools regarding principals' replacement, average teacher characteristics, and the number of longer school days the government implemented in the full-day schools. We found some evidence that the effects of the full-day program on student test scores were larger in schools that did not replace their principals. We did not find strong evidence that heterogeneity in the effects of the program reflected changes in average characteristics of teachers. Notably, schools that replaced more teachers are not associated with larger or smaller program effects.

Finally, the effects of the program on student test scores were reasonable larger in schools that implemented a full-day program 5 days per week (0.39sd in math and 0.26sd in language) than in schools that implemented a 3-days-per-week version of the program (0.20sd in math and 0.18sd in language). We argue that this difference in the effects is associated with the use of time. Whereas the two versions of the program schools increased the hours in core subjects by the same amount, the 5-days-per-week version offered additional supportive activities to enhance academic opportunities. Students in 5-days-per-week full-day schools could participate in supervised independent studies or remedial classes, which, as shown in previous literature (Cortes and Goodman, 2014), has positive impacts on student achievement.

This study is the first paper that provides such detailed evidence on the causal effects of the full-day high school program in Pernambuco, Brazil. The program has caught the attention of international organizations, and the attractiveness has encouraged the spread of the full-day school reform to the national level in Brazil. Moreover, this is one of the few papers that has analyzed the effects of increasing school hours in Latin America. In this context, our study might help to reconcile some of the mixed results found by other studies in Latin America. In general, the changes in school hours analyzed in those studies vary a lot according to the type of full-day school implemented. We show strong evidence that increasing the time of instruction on the core components of the curriculum (math, language, and science) has positive effects on student achievement. Finally, we believe that our results also strengthen the evidence that school instructional time—in particular, the length of the school day—combined with more resources for schools, increases student test scores. One limitation of our work is the absence of a cost-benefit analysis. Future research could investigate and explore what the costs and returns of the program are. Getting information for costs and returns is very challenging in contexts like Brazil and other developing countries; any contribution in this direction would be very informative for decision makers.

References

Agüero, J. M., and T. Beleche. 2013. "Test-Mex: Estimating the effects of school year length on student performance in Mexico." *Journal of Development Economics* 103:353–361.

Alfaro, P., D. K. Evans, and P. Holland. 2015. "Extending the school day in Latin America and the Caribbean." *Policy Research Working Paper* 7309.

Almeida, R., A. Bresolin, B. Borges, K. Mendes, and N. Menezes-Filho. 2016. *Assessing the Impacts of Mais Educação on Educational Outcomes: Evidence between 2007 and 2011.* The World Bank. doi:10.1596/1813-9450-7644. eprint: https://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-7644. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7644.

Angrist, J., E. Bettinger, and M. Kremer. 2006. "Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia." *American economic review* 96 (3): 847–862.

Asim, M., and T. Dee. 2016. *Mobile Phones, Civic Engagement, and School Performance in Pakistan.* Working Paper, Working Paper Series 22764. National Bureau of Economic Research, October. doi:10.3386/w22764. http://www.nber.org/papers/w22764.

Banerjee, A., P. Glewwe, S. Powers, and M. Wasserman. 2013. "Expanding Access and Increasing Student Learning in Post-Primary Education in Developing Countries: A Review of the Evidence."

Battistin, E., and E. C. Meroni. 2016. "Should we increase instruction time in low achieving schools? Evidence from Southern Italy." *Economics of Education Review* 55:39–56.

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

Bettinger, E. P., L. Fox, S. Loeb, and E. S. Taylor. 2017. "Virtual Classrooms: How Online College Courses Affect Student Success." *American Economic Review* 107, no. 9 (September): 2855–75. doi:10.1257/aer.20151193. http://www.aeaweb.org/articles?id=10.1257/aer.20151193.

Bettinger, E. P., and B. T. Long. 2009. "Addressing the needs of underprepared students in higher education does college remediation work?" *Journal of Human resources* 44 (3): 736–771.

Bruns, B., D. Evans, and J. Luque. 2011. *Achieving world-class education in Brazil: The next agenda.* The World Bank.

Cabrera-Hernández, F. 2020. "Does Lengthening the School Day Increase School Value-Added? Evidence from a Mid-Income Country." *The Journal of Development Studies* 56 (2): 314–335.

Cortes, K. E., and J. S. Goodman. 2014. "Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement." *American Economic Review* 104 (5): 400–405.

Guimaraes, P., and R. C. Lindrooth. 2007. "Controlling for overdispersion in grouped conditional logit models: A computationally simple application of Dirichlet-multinomial regression." *The Econometrics Journal* 10 (2): 439–452.

Huebener, M., S. Kuger, and J. Marcus. 2017. "Increased instruction hours and the widening gap in student performance." *Labour Economics* 47:15–34.

INEP. 2019a. *Avaliação Nacional do Rendimento Escolar - Prova Brasil.* Data retrieved from INEP microdata, http://portal.inep.gov.br/.

———. 2019b. *Censo Escolar.* Data retrieved from INEP microdata repository, http://portal.inep.gov.br/.

Lavy, V. 2015. "Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries." *The Economic Journal* 125 (588): F397–F424.

———. 2016. "Expanding School Resources and Increasing Time on Task: Effects on Students' Academic and Noncognitive Outcomes." *Journal of the European Economic Association.*

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

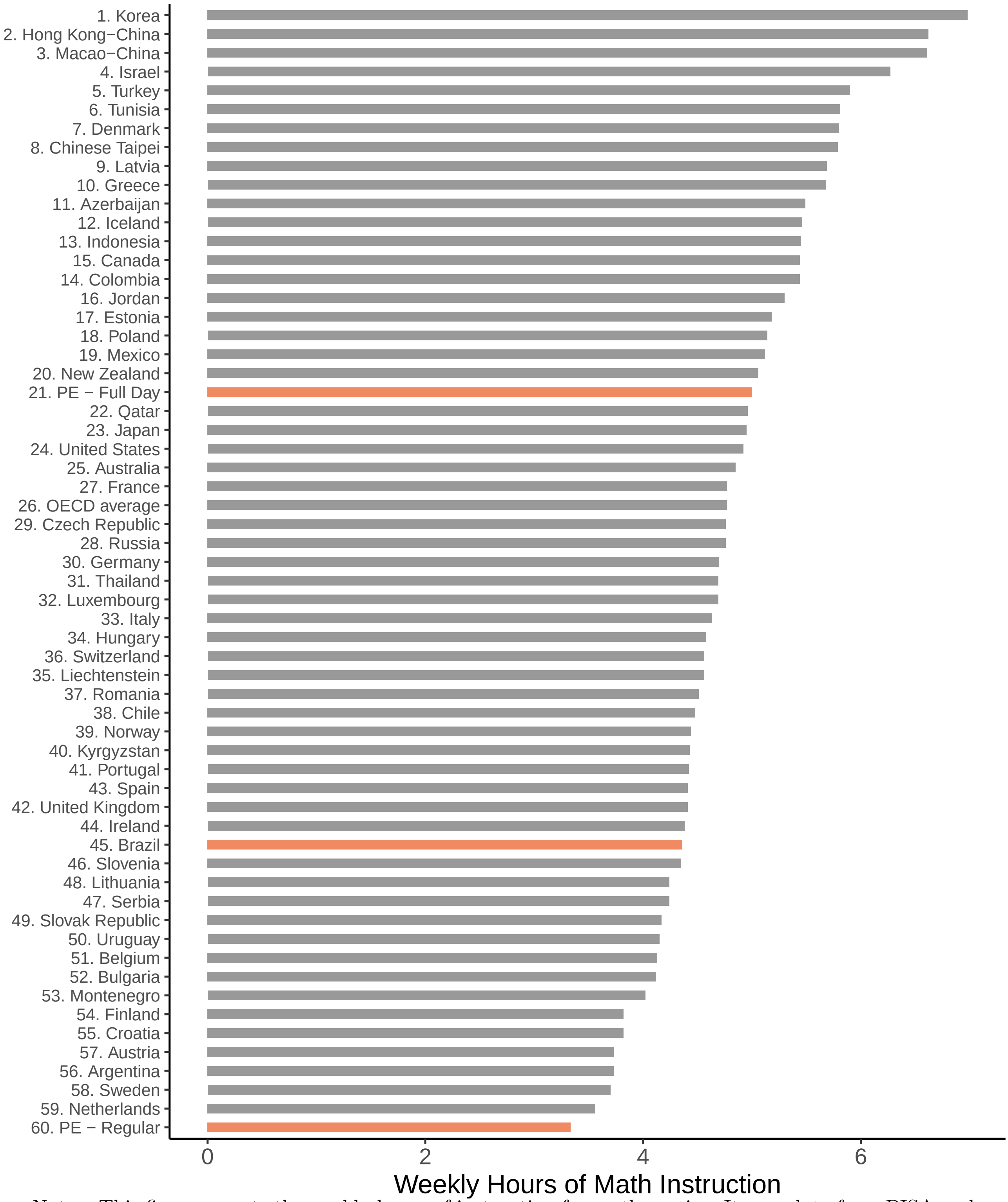
OECD. 2016. *PISA 2015 results (Volume II): Policies and practices for successful schools.*

Pernambuco. 2012. *Diário Oficial 28 de fevereiro de 2012, INSTRUÇÃO NORMATIVA Nº 01/2012.*

Rivkin, S. G., and J. C. Schiman. 2015. "Instruction time, classroom quality, and academic achievement." *The Economic Journal* 125 (588): F425–F448.

Taylor, E. 2014. "Spending more of the school day in math class: Evidence from a regression discontinuity in middle school." *Journal of Public Economics* 117:162–181.

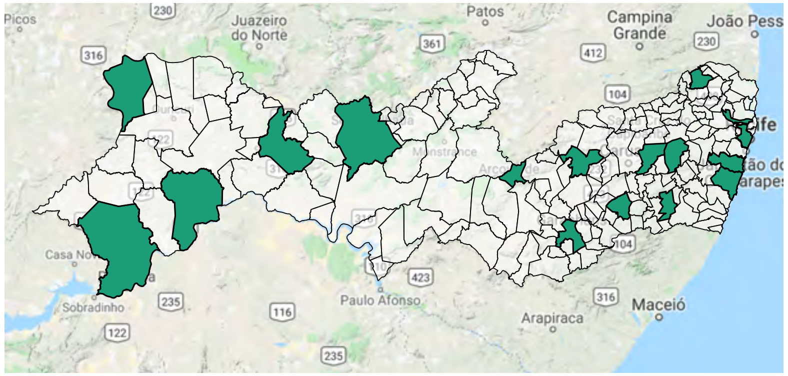
Figure 1
Time of Instruction Across the World - Mathematics



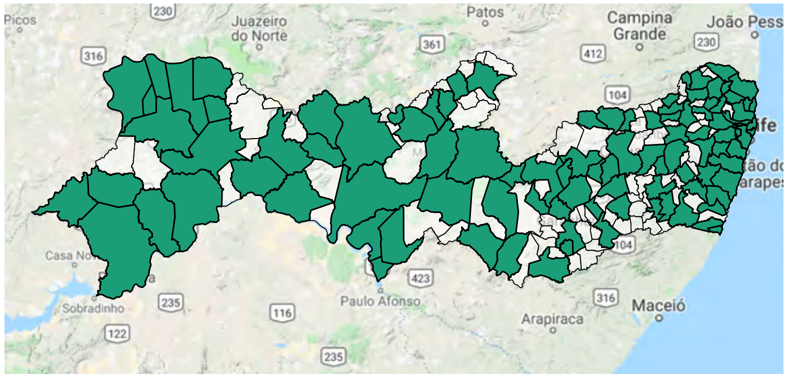
Notes: This figure reports the weekly hours of instruction for mathematics. It uses data from PISA and for all countries (except Pernambuco) it considers time in school and time of instruction out of schools.

Figure 2
Geographical expansion of the full-day high school program

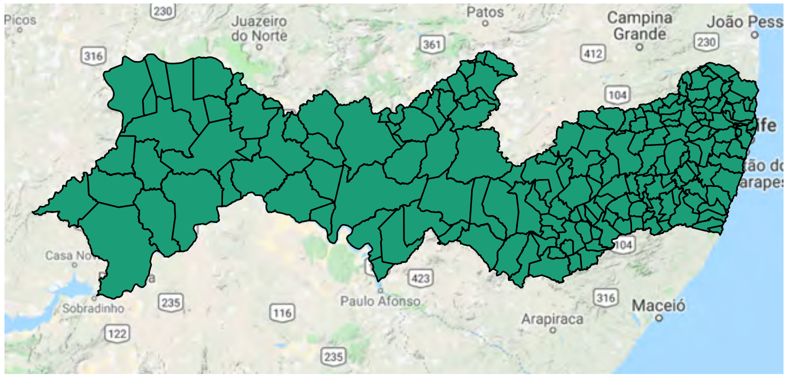
(a) Municipalities with a full-day school in 2004



(b) Municipalities with a full-day school in 2007



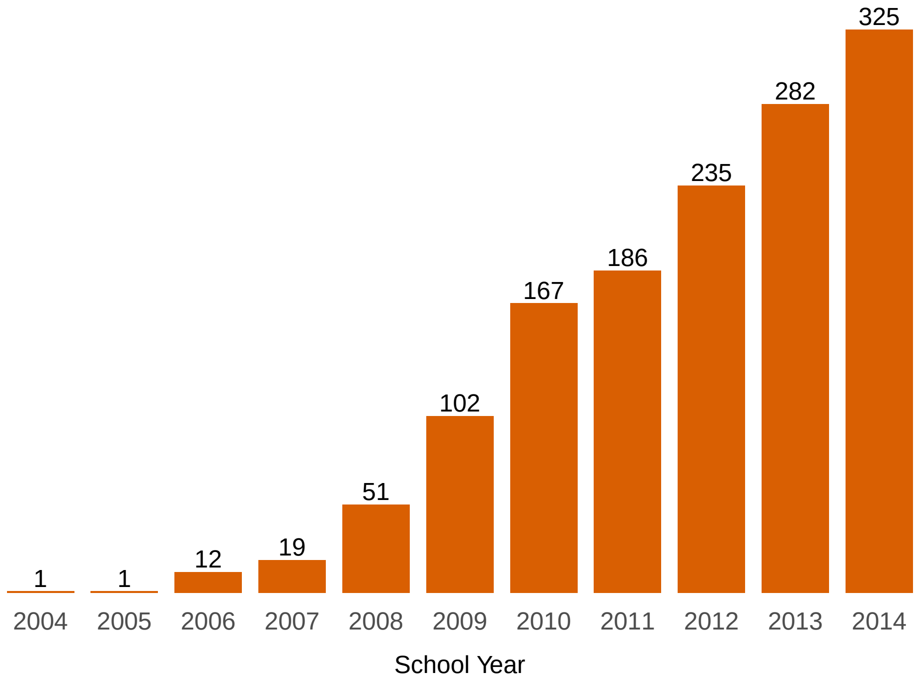
(c) Municipalities with a full-day school in 2010



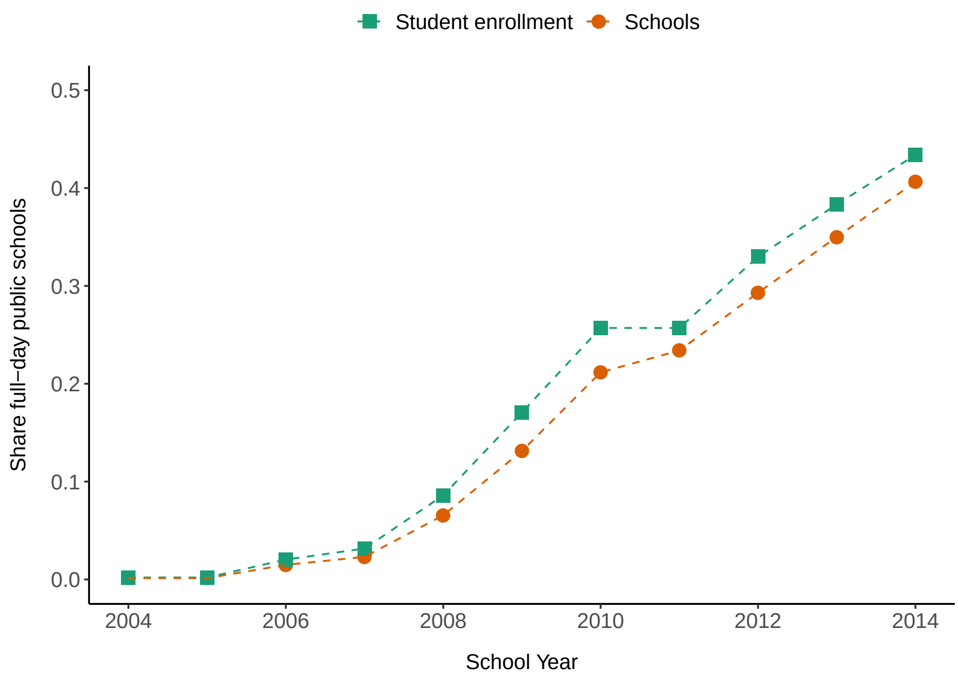
(d) Municipalities with a full-day school in 2014

Notes: These maps report the expansion of the full-day public high school program across Pernambuco's municipalities. Black borders indicate municipality limits. Dark (green) colors indicate the municipalities with at least one full-day school.

Figure 3
Full-day schools: number of schools and participation in the public high school market



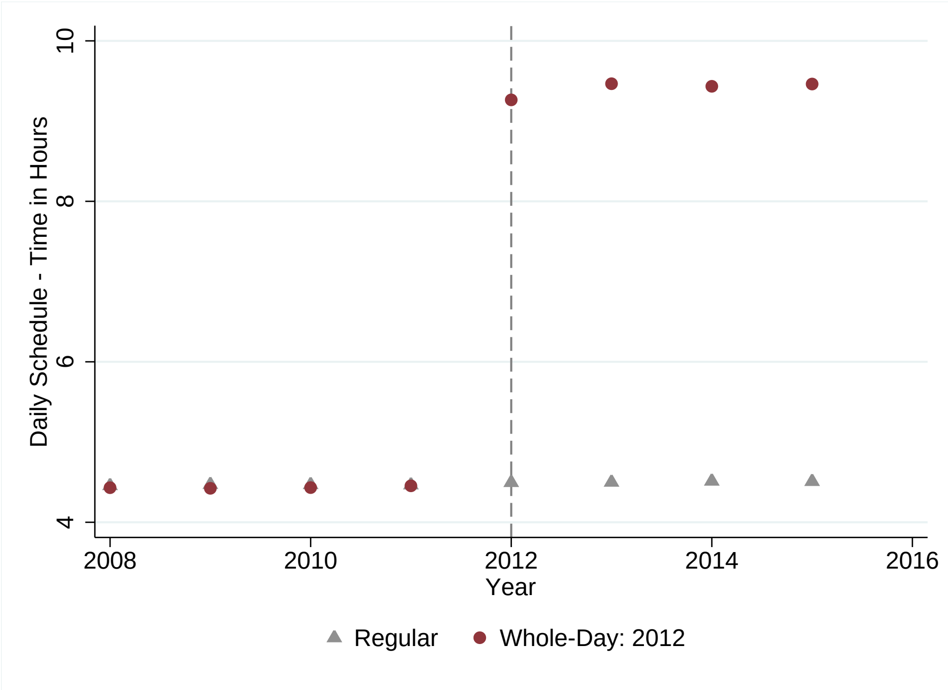
(a) Number of full-day high schools per year



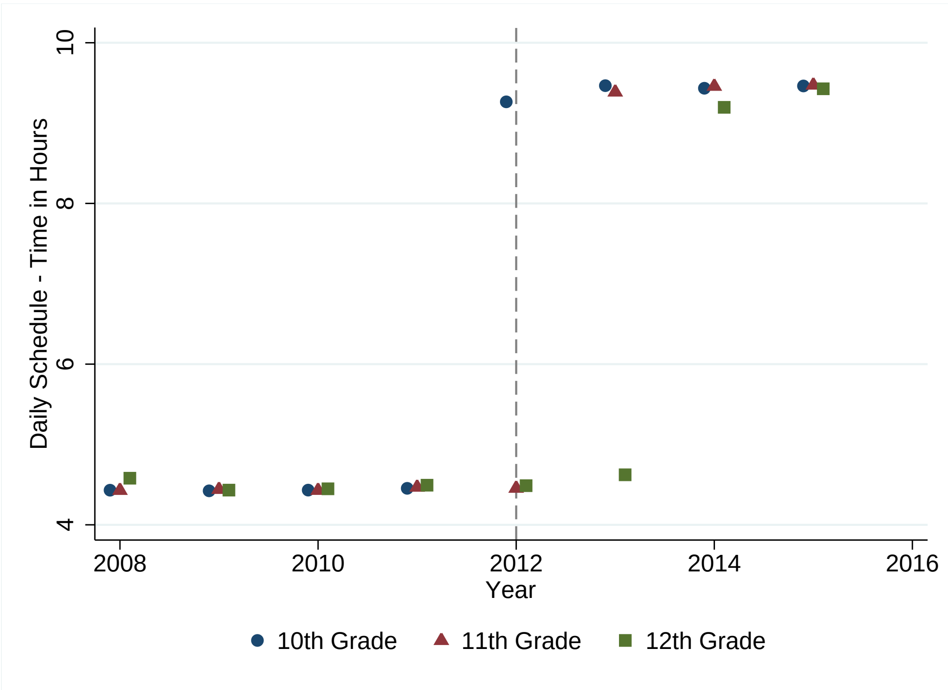
(b) Share of public high school students and buildings in full-day high schools

Notes: These figures report the number of schools converted into full-day high schools per year (top) and the percentage of all public high schools and first-year public high school students in full-day high schools (bottom).

Figure 4
School day hours - Pernambuco



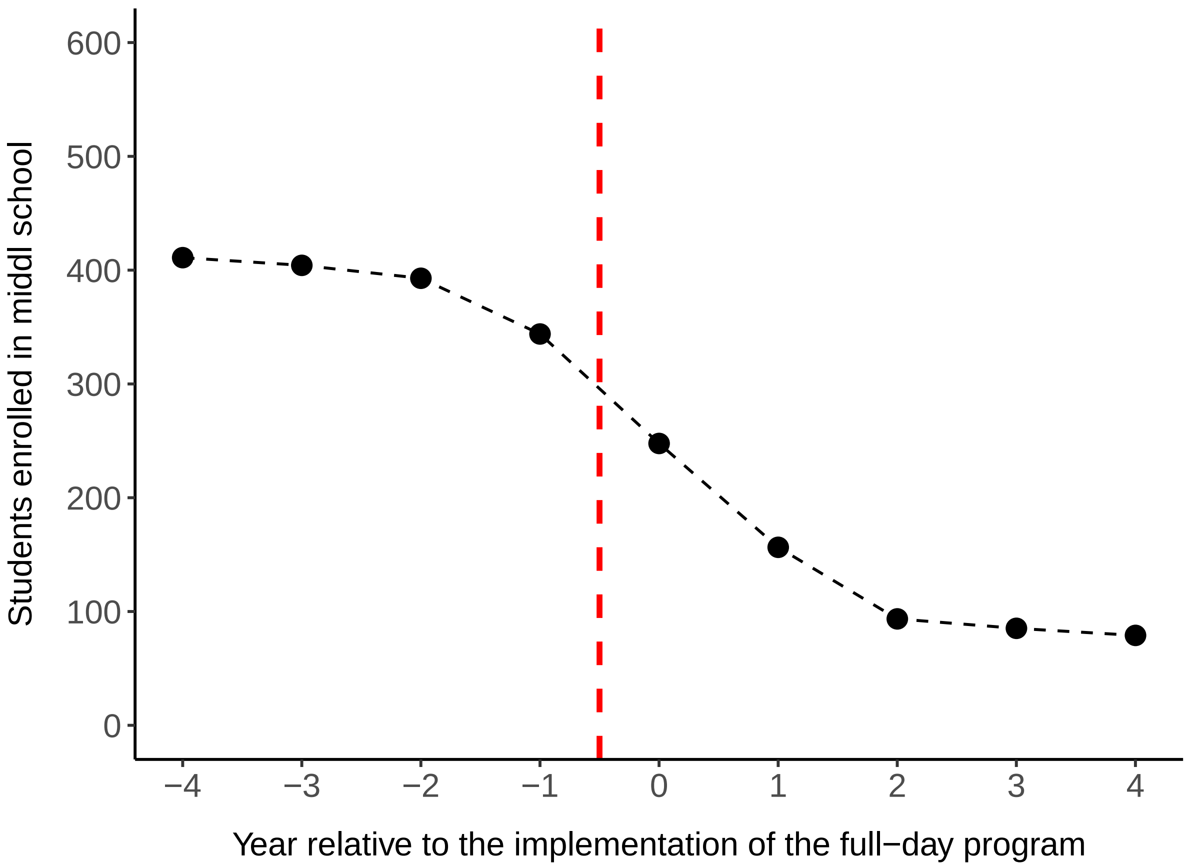
(a) Regular vs full-day schools



(b) Full-day schools by high school grade

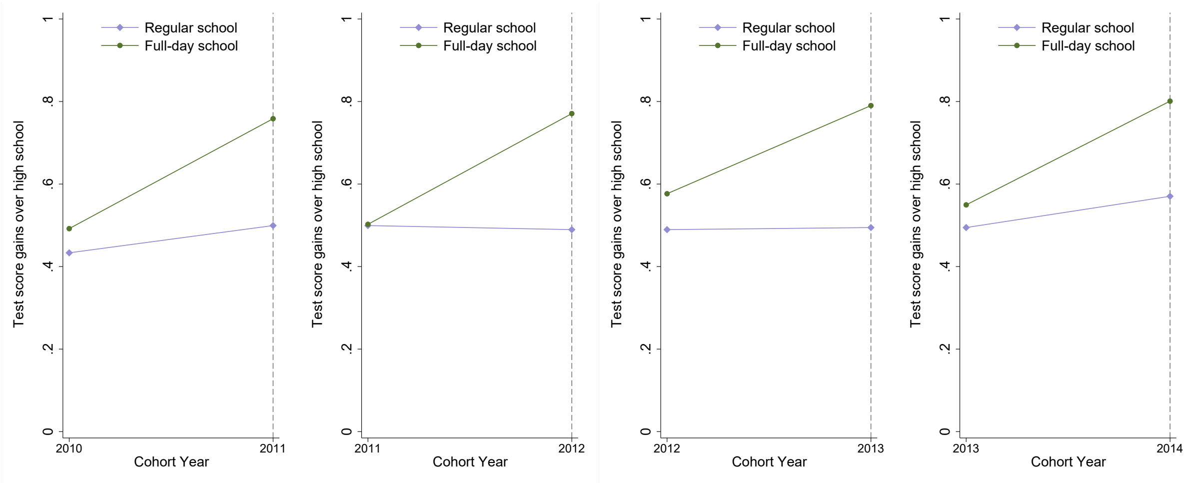
Notes: These figures report the changes in hours in full-day high schools. These figures only use schools that implemented the program in 2012; for details about full-day school implemented in other years, see Figure C.1 in Appendix C. To construct school day hours, we used data from INEP (2019b).

Figure 5
Middle school enrollment in full-day schools



Notes: This figure reports the average number of students enrolled in middle school grades in schools converted into full-day high schools between 2011 and 2014. The horizontal axis indicates the year relative to the implementation of the full-day program. To report these numbers, we used data from INEP (2019b).

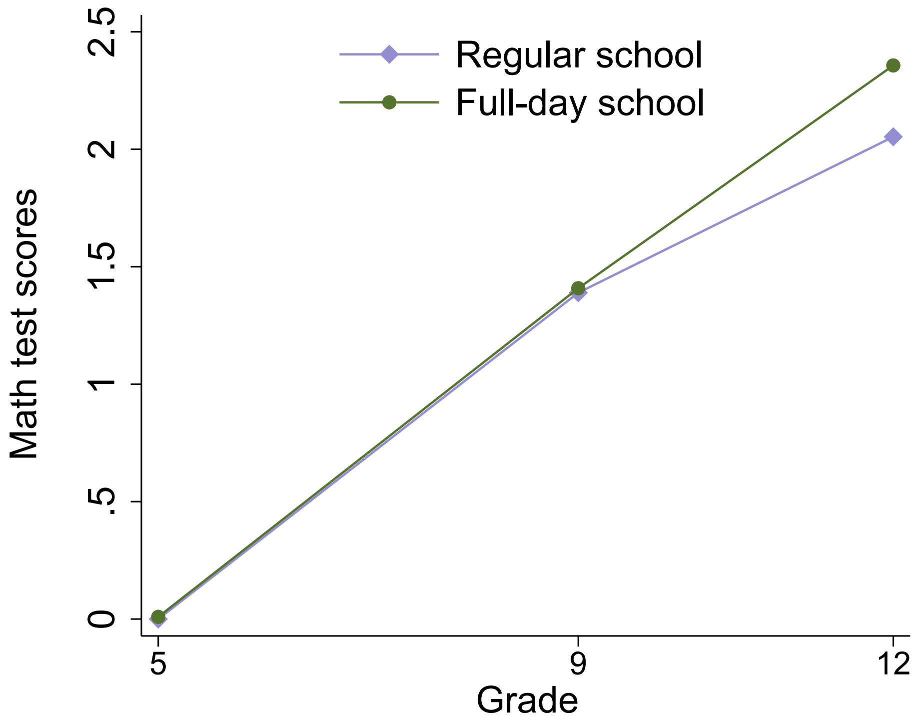
Figure 6
High school math test score gains by cohort and school type



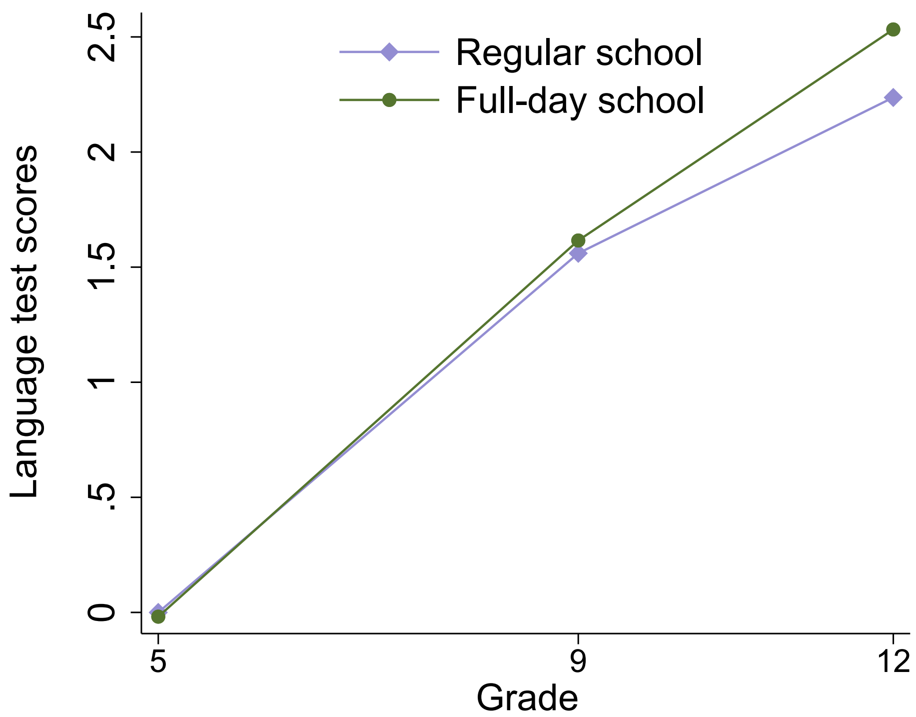
(a) Full-day schools converted in 2011 (b) Full-day schools converted in 2012 (c) Full-day schools converted in 2013 (d) Full-day schools converted in 2014

Notes: These figures reports math test score gains over high school for students attending regular, full-day schools before the program, and full-day schools after the program (treatment group). Each figure reports gains for a different group of full-day high schools, i.e., schools that implemented the full-day program in different years.

Figure 7
Difference-in-differences comparing student test scores from elementary school to high school (high school cohort of 2014)



(a) Mathematics



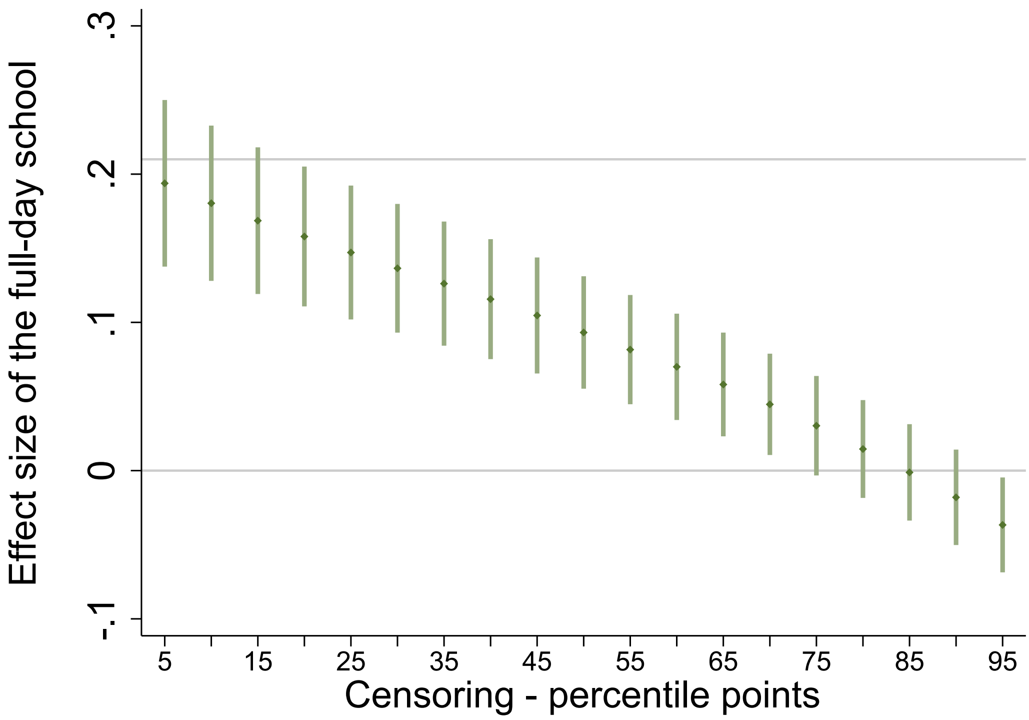
(b) Language

Notes: These figures report (unconditional) average test scores of students in elementary, middle and high school. Because of data availability, it only uses students who started high school in 2014. For a formal test of these differences, please check Table D.4 in Appendix D.

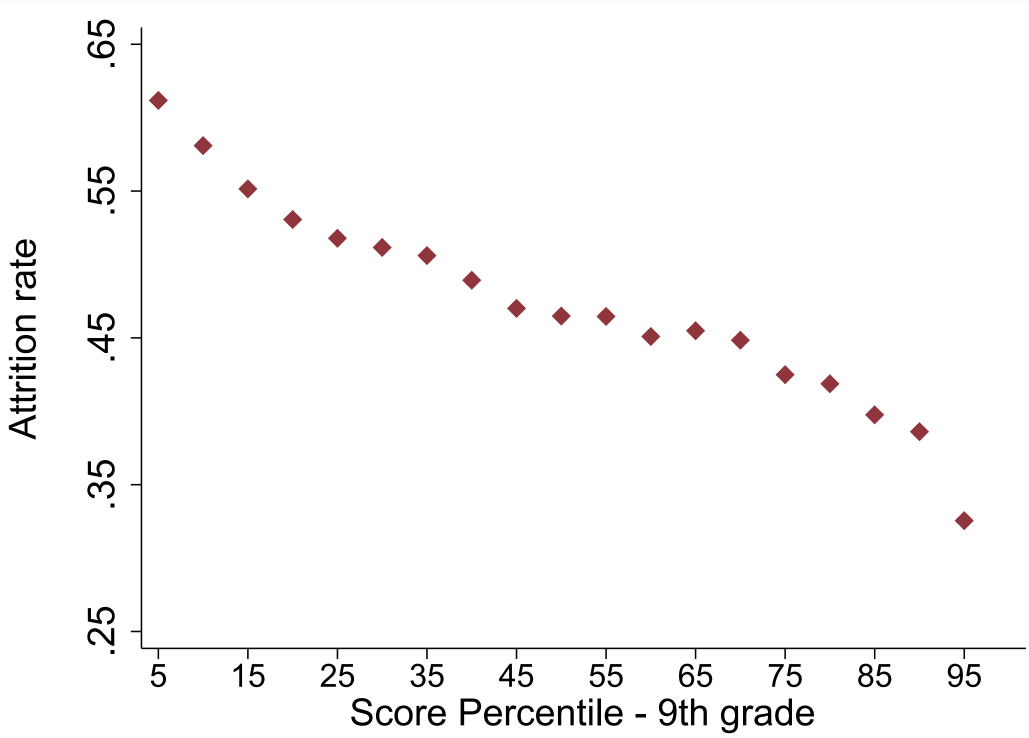
Figure 8

Censored Regressions

Panel A. Effects Size: Censored Regression



Panel B. Attrition rates by student initial test scores (before high school)



Notes: Panel A reports the effects of the full-day school program for different censored points (as described in section 5: robustness checks). Each point in Panel A is an estimate from a different regression. Percentile points are the points we censored our data. When censoring the data, we assigned all missing values and test scores below these percentiles to the exact percentile point. All results have a total of observations equals 394,736 and include all students we have 9th-grade test scores. Panel B shows the attrition rates according to the percentile scores in the 9th grade (baseline), i.e., before high school.

Table 1
High school timetables (Hours Per Week)

	Regular	Full-Day	
		3-days	5-days
Core			
Humanities			
Language (portuguese)	4.2	5.0	5.0
Social Sciences	5.0	5.0	5.0
STEM			
Mathematics	3.3	5.0	5.0
Sciences	5.3	8.1	8.1
Other			
Foreign Language	1.7	1.7	1.7
Arts	0.6	1.1	1.1
Physical Education	0.8	1.7	1.7
Extra class activities			
"Life-project" program	0	1.7	1.7
Independent/remedial class	0	0	8.3

Notes: This table reports the number of hours in the curriculum of high schools in Pernambuco. All data was collected from public documents (Pernambuco, 2012). For more details, see text in Section 2.3.

Table 2
Principal characteristics

	Regular	Full-Day High Schools				
		Before converted		After converted		
	Mean	Mean	P-value	Mean	P-value	P-value
	(1)	(2)	(2)-(1)	(3)	(3)-(1)	(3)-(2)
Demographics						
Age	47.80	48.59	0.38	47.93	0.88	0.57
Female	0.73	0.80	0.19	0.75	0.69	0.48
Experience, contract, and wage						
Experience: as principal	7.93	9.37	0.03	7.39	0.39	0.02
Experience: as principal at the current school	6.82	7.57	0.23	4.83	0.00	0.00
Experience: in education	15.15	15.43	0.34	15.31	0.56	0.76
Recruitment: Competition	0.03	0.23	0.00	0.67	0.00	0.00
Recruitment: Competition and Election	0.68	0.45	0.00	0.16	0.00	0.00
Recruitment: Election	0.03	0.05	0.41	0.00	0.16	0.10
Recruitment: Indication (technical or political)	0.12	0.18	0.12	0.00	0.00	0.00
Recruitment: other	0.14	0.09	0.27	0.17	0.50	0.19
Wage	2975.15	2639.60	0.09	4085.90	0.00	0.00
Weekly hour workload at the current school	38.69	39.39	0.10	39.72	0.01	0.57

Notes: This table reports characteristics of high school principals using data from the Prova Brasil and SAEB survey data in the years 2011, 2013, and 2015 (INEP, 2019a). The survey is applied every two years, and the principals self-report all information. To have information before and after the program was implemented, we only used schools that implemented the full-day school program between 2012 and 2014. Also, we restricted data for schools that appeared in all surveys. P-values were computed using a t-test.

Table 3
Teacher characteristics

	Regular	Full-day		F-test for equality
		Pre-program	Post-program	(prob > F)
	(1)	(2)	(3)	(4)
Female	0.657	0.662	0.646	0.096
	[0.475]	[0.473]	[0.478]	
Age	40.244	38.851	38.218	0.000
	[10.245]	[10.144]	[9.664]	
Graduated at a state (elite) college	0.659	0.671	0.697	0.104
	[0.474]	[0.470]	[0.460]	
New Teacher at School	0.376	0.328	0.329	0.014
	[0.484]	[0.470]	[0.470]	
Teaching more than a cohort at school	0.891	0.514	0.745	0.000
	[0.311]	[0.500]	[0.436]	
Number of Classrooms	10.711	9.913	9.141	0.038
	[5.643]	[5.271]	[4.803]	
Number of Schools	1.499	1.465	1.408	0.313
	[0.664]	[0.621]	[0.576]	

Notes: This table reports characteristics of high school teacher using data from Censo Escolar INEP, 2019b. The data is collected every year and all information was constructed by identifying teachers and connecting them to high school cohorts.

Table 4
School Characteristics

	Regular	Full-Day High Schools				
		Before converted		After converted		
	Mean	Mean	P-value	Mean	P-value	P-value
	(1)	(2)	(2)-(1)	(3)	(3)-(1)	(3)-(2)
Building						
Classroom						
Quality Index	2.52	2.59	0.38	2.62	0.16	0.72
Good ventilation	0.72	0.81	0.14	0.81	0.10	0.99
Good lighting	0.92	0.91	0.73	0.96	0.30	0.32
Facilities						
Auditorium						
Own	0.36	0.42	0.32	0.45	0.11	0.69
Quality Index	2.48	2.54	0.68	2.59	0.37	0.74
Lab: Arts						
Own	0.07	0.06	0.71	0.14	0.03	0.07
Quality Index	2.47	2.00	0.21	2.00	0.06	1.00
Lab: Music						
Own	0.12	0.07	0.22	0.23	0.02	0.01
Quality Index	2.42	2.40	0.96	2.19	0.26	0.58
Lab: Science						
Own	0.44	0.60	0.01	0.46	0.71	0.10
Quality Index	2.46	2.39	0.56	2.42	0.79	0.84
Library						
Own	0.96	0.99	0.32	0.99	0.29	0.98
Quality Index	2.57	2.57	0.96	2.60	0.72	0.76
Room for group study						
Own	0.65	0.64	0.96	0.84	0.00	0.01
Quality Index	1.00	1.00	0.76	1.00	0.72	1.00
Sport court						
Own	0.69	0.64	0.34	0.61	0.13	0.70
Quality Index	1.99	2.21	0.09	2.00	0.93	0.23

Notes: This table reports characteristics of high schools using data from the Prova Brasil and SAEB survey data in the years 2011, 2013, and 2015 (INEP, 2019a). The survey is applied every two years, and all information is reported by enumerators that visit the school to apply the survey. To have information before and after the program was implemented, we only used schools that implemented the full-day school program between 2012 and 2014. Also, we restricted data for schools that appeared in all surveys. P-values were computed using a t-test.

Table 5
Summary Statistics

	Mean (1)	*Std. Dev.* (2)	*Observations* (3)
Panel A. Student/Household data - Pre 10th grade			
Student			
Math Test Score (SD units)	0.16	0.99	103,812
Age (years)	15.89	1.81	103,812
Female	0.60	0.49	103,812
Racial Minority	0.66	0.47	100,457
Household			
Mother completed secondary	0.25	0.43	100,313
CCT eligible	0.73	0.45	100,604
Computer in the household	0.83	0.37	100,149
Less than 2 bathrooms	0.39	0.49	100,181
Less than 20 books at home	0.78	0.41	100,502
Panel B. Student-Post 10th grade			
Math Test Score in 12th grade	0.69	1.11	103,812
Average gain between 9th and 12th grade	0.53	1.01	103,812

Note: All statistics reported are from our restricted sample, as described in Section 3. This sample includes only students who have non-missing test scores in 9th and 12th grades. Student data are from the administrative records of Pernambuco, Brazil. Student and household characteristics are measured in 9th grade and are self-reported by students in a survey applied after the state tests.

Table 6
Student characteristics by regular and full-day schools

	Regular (Mean)	Full-day Schools		F-test for equality
		Pre-program	Post-program	
	(1)	(2)	(3)	(4)
Student Characteristics				
Age (years)	15.929	15.894	15.617	0.0000
	[1.883]	[1.741]	[1.177]	
Female	0.592	0.618	0.613	0.3089
	[0.491]	[0.486]	[0.487]	
Race: Black or Brown	0.670	0.662	0.666	0.1848
	[0.470]	[0.473]	[0.472]	
Household Characteristics				
Mother completed secondary education	0.245	0.218	0.236	0.1342
	[0.430]	[0.413]	[0.425]	
CCT eligible	0.726	0.771	0.765	0.4936
	[0.446]	[0.420]	[0.424]	
Computer in the household	0.387	0.293	0.385	0.1674
	[0.487]	[0.455]	[0.487]	
Less than 2 bathrooms	0.837	0.845	0.830	0.1962
	[0.369]	[0.361]	[0.376]	
Less than 20 books	0.786	0.798	0.812	0.4751
	[0.410]	[0.402]	[0.391]	
Student test scores				
Math Tests Score in 9th grade	0.144	0.152	0.280	0.0140
	[0.999]	[0.963]	[0.963]	
Math Test Score in 12th grade	0.645	0.690	1.063	0.0000
	[1.099]	[1.060]	[1.162]	
Reading Test Score in 9th grade	0.172	0.142	0.307	0.0035
	[1.053]	[1.012]	[1.053]	
Reading Score in 12th grade	0.786	0.811	1.166	0.0000
	[1.152]	[1.135]	[1.114]	

Notes: This table reports summary statistics for regular and full-day schools. Column 4 shows the p-value for an F-test that compares students in full-day schools after the program started with other students, after accounting for year and municipality fixed effects. Standard deviations are reported in square brackets.

Table 7
Triple Differences Model (non-parametric)

Panel A: Treatment and Control students attending full-day school buildings

School	High school	Test Scores			Diff-in-diff
Type	Cohort	Grade 9	Grade 12	High school gains	
Full-day	t	$Y^F_{g=9,t}$	$Y^F_{g=12,t}$	$\Delta Y^F_t = Y^F_{g=12,t} - Y^F_{g=9,t}$	
Full-day	$t-1$	$Y^F_{g=9,t-1}$	$Y^F_{g=12,t-1}$	$\Delta Y^F_{t-1} = Y^F_{g=12,t-1} - Y^F_{g=9,t-1}$	$\Delta Y^F_t - \Delta Y^F_{t-1}$

Panel B: Treatment and Control students attending regular school buildings

School	High school (HS)	Test Scores			Diff-in-diff
Type	Cohort	Grade 9	Grade 12	HS gains	
Regular	t	$Y^R_{g=9,t}$	$Y^R_{g=12,t}$	$\Delta Y^R_t = Y^R_{g=12,t} - Y^R_{g=9,t}$	
Regular	$t-1$	$Y^R_{g=9,t-1}$	$Y^R_{g=12,t-1}$	$\Delta Y^R_{t-1} = Y^R_{g=12,t-1} - Y^R_{g=9,t-1}$	$\Delta Y^R_t - \Delta Y^R_{t-1}$

Panel C: Triple differences estimator

Full-day program effects $= (\Delta Y^F_t - \Delta Y^F_{t-1}) - (\Delta Y^R_t - \Delta Y^R_{t-1})$

Table 8

Triple Differences Model: Effects of full-day high schools on student test scores

	Math				Language
	(1)	(2)	(3)	(4)	(5)
Full-day × Grade$_{g=12}$ × Cohort$_{t=t^*}$	0.220***	0.220***	0.220***	0.225***	0.195***
	(0.064)	(0.042)	(0.041)	(0.015)	(0.012)
Full-day × $Grade_{g=12}$	0.048	0.048	0.048		
	(0.047)	(0.030)	(0.028)		
Full-day ×Cohort$_{t=t^*}$	0.129**	0.098**	0.063		
	(0.042)	(0.034)	(0.033)		
Full-day	0.084	0.004	-0.016		
	(0.052)	(0.070)	(0.064)		
$Grade_{g=12}$	0.434***	0.434***	0.434***		
	(0.032)	(0.025)	(0.023)		
Municipality Fixed Effects	No	Yes	Yes	-	-
Student Characteristics	No	No	Yes	Yes	Yes
School FE × Cohort	-	-	-	Yes	Yes
School FE × Grade	-	-	-	Yes	Yes
Observations	207,624	207,624	207,624	207,624	207,615

Notes: This table reports the effects of the full-day high school on student test scores. The first is the interaction of (Full-day$_j$ × Grade$_{g=12}$ × Cohort$_{t=t^*}$) as shown in Equation 1 and reports the effects of the full-day high school program on student test scores. All columns controls for $Cohort_t$ and $Cohort_t \times$ Full-day. Columns 4 and 5 replace the auxiliary fixed effects described in Equation 1 by interactions with school fixed effects. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 9
Triple Differences Model: Effects of full-day high schools on student test scores

	Year the full-day program started			
	2011	2012	2013	2014
	(1)	(2)	(3)	(4)
Panel A. Mathematics test scores				
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.222***	0.264***	0.212***	0.195***
	(0.0387)	(0.0263)	(0.0292)	(0.0346)
F-test for equality (prob > F)	.396			
N	65,704	80,966	81,622	81,406
Panel B. Language test scores				
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.273***	0.211***	0.194***	0.146***
	(0.0309)	(0.0221)	(0.0229)	(0.0239)
F-test for equality (prob > F)	.043			
N	65,700	80,961	81,622	81,402

Notes: This table reports the effects of the full-day high school on student test scores by year of implementation of the full-day high school program. All columns control for $school \times Grade$, $school \times Cohort$, $Cohort \times Grade$, and student characteristics. For more details, see notes in Table 8. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 10

Effects of full-day high schools on student test scores using the instrumental variable approach

	Triple Differences	IV1: P(Attend FD) in 2008		IV2: Treated Middle School		IV3: IV2 and IV3	
		1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Math test scores							
Full-day $\times$ Grade$_{g=12}$ $\times$ Cohort$_{t=t^*}$	0.220***		0.307*		0.070		0.228*
	(0.063)		(0.145)		(0.129)		(0.113)
P(Attend FD) in 2008		0.291***				0.239***	
		(0.015)				(0.016)	
Treated Middle School				0.283***		0.181***	
				(0.019)		(0.017)	
F (weak test)			358		227		209
Hasen Test: p-value							0.174
N	207,624	207,624	207,624	207,624	207,624	207,624	207,624
Panel B. Language test scores							
Full-day $\times$ Grade$_{g=12}$ $\times$ Cohort$_{t=t^*}$	0.194***		0.169		0.193		0.177*
	(0.051)		(0.110)		(0.102)		(0.088)
P(Attend FD) in 2008		0.291***				0.239***	
		(0.015)				(0.016)	
Treated Middle School				0.283***		0.181***	
				(0.019)		(0.017)	
F (weak test)			358		227		209
Hasen Test: p-value							0.854
N	207,615	207,615	207,615	207,615	207,615	207,615	207,615

Notes: This table reports the effects of the full-day high school on student test scores. Detailed description of the instruments are presented in section 4. Columns 2, 4, and 6 report the first state relationship between our treatment dummy (Full-day $\times$ Grade$_{g=12}$ $\times$ Cohort$_{t=t^*}$) and our instruments. Columns 3, 5, and 7 shows the full-day program effects estimates using the different instruments. The Hansen test reported in Column 7 is an exclusion test that checks if one of the instruments is invalid. In all models, we used as the endogenous model the specification described in equation 1. For alternative specifications that include municipality fixed effects or school fixed effects, see Table D.2 and D.3 in Appendix D. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 11
Triple differences: Effects of the full-day program on students' composition

	Age	Female	Black	Mother's Education
	(1)	(2)	(3)	(4)
Full-day × Grade$_{g=12}$ × Cohort$_{t=t^{*}}$	-0.009	-0.003	0.007	0.002
	(0.006)	(0.001)	(0.005)	(0.018)
N	207,624	207,624	183,091	159,512
	Number of Books	Has a Computer	Bathrooms	CCT
	(5)	(6)	(7)	(8)
Full-day × Grade$_{g=12}$ × Cohort$_{t=t^{*}}$	0.007	0.051***	-0.014	0.010
	(0.013)	(0.009)	(0.008)	(0.008)
N	182,841	182,464	181,796	182,599

Table 12

Heterogeneous effects of full-day high school program on student math test scores by status of principal replacement at the full-day schools

	Principal replacement in the full-day school			
	No re-placement	Before the program	At the year the program started	After the program started
	(1)	(2)	(3)	(4)
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.310^{***}	0.175^{***}	0.201^{***}	0.221^{***}
	(0.0316)	(0.0401)	(0.0207)	(0.0310)
F-test for equality (prob > F)	0.0166			
N	177,008	175,654	181,520	172,846

Notes: This table reports the estimates of heterogeneous effects of the full-day high school program on student test scores. It divides the full-day high schools into four groups. Column 1 reports the effects of the program for full-day schools that did not change the principal from 2009 to 2015. Column 2 reports the effects of full-day schools that changed the principal before the program started. Column 3 reports the effects for full-day schools that changed their principal in the year the program started. Finally, column 4 reports the effects of full-day schools that changed their principal one or two years after the program started. The row "F-test for equality" reports the p-value for a test of equality across Columns 1-4. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 13

Heterogeneous effects of full-day high school program on student math test scores by teacher average characteristics of full-day schools

	Female		Elite college		New Teacher	
	Below median	Above median	Below median	Above median	Below median	Above median
	(1)	(2)	(3)	(4)	(5)	(6)
Full-day × $Grade_{g=12}$ × $Cohort_{t=t^*}$	0.275^{***}	0.177^{***}	0.210^{***}	0.240^{***}	0.217^{***}	0.234^{***}
	(0.0232)	(0.0195)	(0.0198)	(0.0231)	(0.0231)	(0.0203)
F-test for equality (prob > F)		0.001		0.343		0.566
N	186,438	187,654	187,126	186,966	186,164	187,928
	Teaching more than one cohort		Number of classrooms		Number of schools	
	Below median	Above median	Below median	Above median	Below median	Above median
	(7)	(8)	(9)	(10)	(11)	(12)
Full-day × $Grade_{g=12}$ × $Cohort_{t=t^*}$	0.226^{***}	0.225^{***}	0.191^{***}	0.253^{***}	0.233^{***}	0.219^{***}
	(0.0203)	(0.0226)	(0.0249)	(0.0188)	(0.0224)	(0.0210)
F-test for equality (prob > F)		0.9256		0.0378		0.6193
N	186,470	187,622	184,544	189,548	186,066	188,026

Notes:This table reports estimates of the effects of the full-day high school program by teacher average characteristics teaching at full-day high schools. For each average characteristics we divided schools in two groups: below the median and above the median. The columns low and high make references to these groups, respectively. All columns control for *school* × *Grade*, *school* × *Cohort*, *Cohort* × *Grade*, and student characteristics. For more details, see notes in Table 8. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 14
Heterogeneous effects of full-day high school program on student math test scores by days of exposure to the full-day program

	Math		Language	
	3 days of full-day program	5 days of full-day program	3 days of full-day program	5 days of full-day program
	(1)	(2)	(3)	(4)
Full-day $\times$ Grade$_{g=12}$ $\times$ Cohort$_{t=t^*}$	0.196*** (0.0167)	0.397*** (0.0317)	0.184*** (0.0131)	0.257*** (0.0234)
F-test for equality (prob > F)		0.0000		0.0045
N	201,540	172,552	201,531	172,545

Notes:This table reports estimates of the effects of the full-day high school program by days of exposure to the full-day program. Full-day high schools might be implemented as 3 days or 5 days of extended school hours. For more details about the different full-day high schools, see Section 2. All columns control for $school \times Grade$, $school \times Cohort$, $Cohort \times Grade$, and student characteristics. For more details, see notes in Table 8. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Appendix A Analytical Sample Construction

Section 3 describes briefly how we defined our analytical sample. This appendix describes in detail the process we followed to reach that definition.

Using enrollment administrative records, we selected all students enrolled in 10th grade in one of the following years: 2010, 2011, 2012, 2013 and 2014. We follow these students to find their records before the 10th grade, and at least two calendar years later, when they might have reached 12th grade.

Subsequently, we performed a series of exclusions. We began by excluding students not enrolled in state schools (mainly students enrolled in private schools). Next, we dropped students in vocational high-schools, and schools that implemented the full-day program before 2011. Finally, within schools, we dropped students for whom we could not find enrollment records in 9th grade. Furthermore, for schools that implemented the full-day school program starting in 2011, we excluded cohorts of students that were enrolled in 10th grade for two or more years before the school began the program, and one or more years after the school started the program.

Using this already reduced data, we checked if a school had more than ten students per year in the 10th grade cohort and achievement gains for all 10th-grade cohorts. If a school did not have at least ten students enrolled in the 10th-grade or did not show gains for one the cohorts, we excluded the school.

Then, we imposed some final restrictions. First, we excluded students for whom we could not find test scores in both 9th and 12th grades. Second, we excluded one school that started the program in 2014. The State Department of Education informed us that this specific school did not follow the implementation policy.

Appendix B Student Matching

We use two primary datasets: enrollment records and test score data. Although these files do not share a common identifier, they have some variables in common that allow us to merge them. We used three main variables in the merging:

1. Name of student
2. Grade in which the student is enrolled (9th or 12th grade).
3. School of student's enrollment

Combining these three variables, we searched for students within the school within the grade, which reduces duplicated names and false positives dramatically. We applied the following procedure in the search process:

- Step A: Drop duplicated names within school and grade in each dataset.
- Step B: Merge both datasets using full-name, grade, and school.

After Step B was completed, we had two groups. The first group was a group of students that we find in the merging process, which we defined as matched. The rest we defined as non-matched and follow for a next step.

- Step C: Merge both datasets using first name, second last name, last name, grade, and school.

We repeated the procedure in step B, creating two groups. The first group, the group of matched students, was appended in the group matched in step B. The rest we defined as non-matched and tried in the last round to see if we could match them.

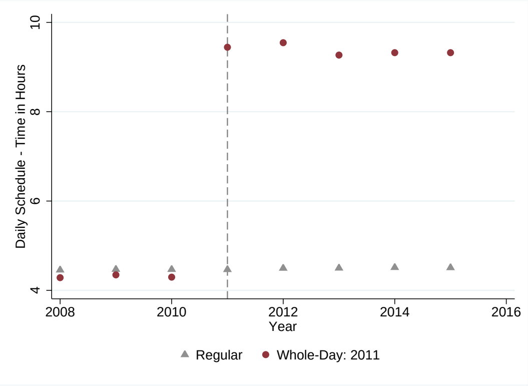
- Step D: Merge both datasets using the full name of students, grade, and school. But now applying a string matching algorithm.

We used two measures in our string matching: Levenshtein, and Jaro–Winkler measures. We were conservative when picking the best matches predicted by these measures, and excluded all matches that had values lower than 4 for the Levenshtein measure or Jaro-Winkler larger than 0.85.

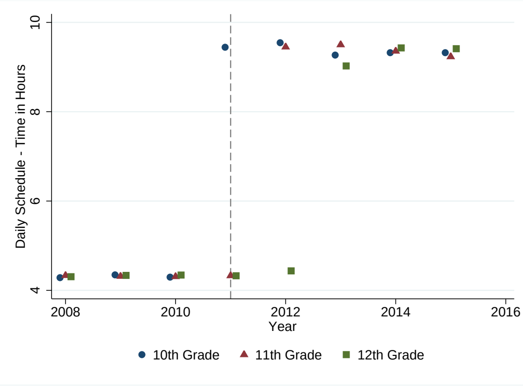
Using this process, we found about 90-95% of students tested (conditional on having a name) in most of years. The only exception is the year 2016 where we found about 83% of the students tested.

Appendix C Figures

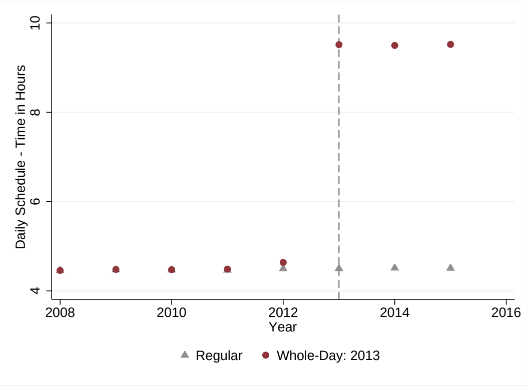
Figure C.1
School day hours - Pernambuco



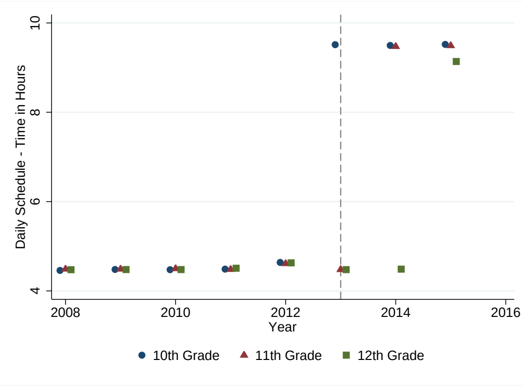
(a) Regular vs full-day schools (2011)



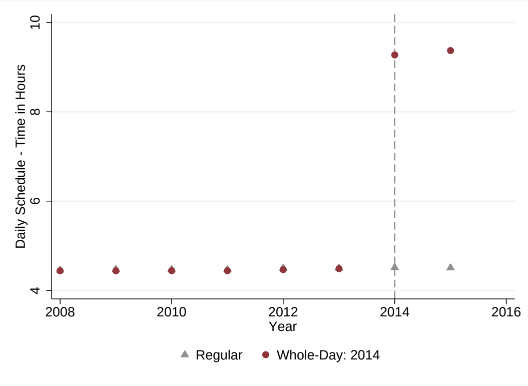
(b) Full-day schools by high grade (2011)



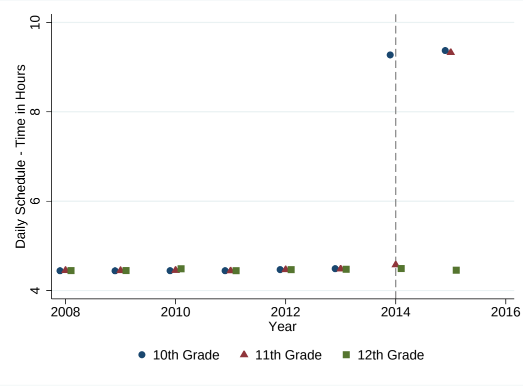
(c) Regular vs full-day schools (2013)



(d) Full-day schools by high grade (2013)



(e) Regular vs full-day schools (2014)



(f) Full-day schools by high grade (2014)

Appendix D Tables

Table D.1
Example of dataset to estimate the group conditional logit

Group ID	Middle Sch (MS) ID	High Sch (HS) ID	Distance MS-HS	Private HS	Size HS	Students from MS to HS
1	MS1	HS1	D11	P1	S1	N11
1	MS1	HS2	D12	P2	S2	N12
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	MS2	HS1	D21	P1	S1	N21
2	MS2	HS2	D22	P2	S2	N22
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	MSK	HSK	DK1	P1	S1	NK1
K	MSK	HSK	DK2	P2	S2	NK2
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table D.2
Triple Differences Model: Effects of full-day high schools on student test scores
Specification including municipality fixed effects

	Triple Differences	IV1: P(Attend FD) in 2008		IV2: Treated Middle School		IV3: IV2 and IV3	
		1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Math test scores							
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.220***		0.302***		0.183*		0.267***
	(0.041)		(0.088)		(0.093)		(0.073)
P(Attend FD) in 2008		0.327***				0.270***	
		(0.016)				(0.016)	
Treated Middle School				0.293***		0.174***	
				(0.018)		(0.017)	
F (weak test)			444		260		244
Hasen Test: p-value							0.290
N	207,624	207,624	207,624	207,624	207,624	207,624	207,624
Panel B. Language test scores							
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.194***		0.266***		0.227**		0.255***
	(0.034)		(0.074)		(0.082)		(0.062)
P(Attend FD) in 2008		0.327***				0.270***	
		(0.016)				(0.016)	
Treated Middle School				0.293***		0.174***	
				(0.018)		(0.017)	
F (weak test)			444		260		244
Hasen Test: p-value							0.683
N	207,615	207,615	207,615	207,615	207,615	207,615	207,615

Table D.3
Triple Differences Model: Effects of full-day high schools on student test scores
Specification including school fixed effects

	Triple Differences	IV1: P(Attend FD) in 2008		IV2: Treated Middle School		IV3: IV2 and IV3	
		1st Stage	2SLS	1st Stage	2SLS	1st Stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Math test scores							
Full-day $\times$ Grade$_{g=12}$ $\times$ Cohort$_{t=t^*}$	0.225***		0.156***		0.270***		0.179***
	(0.015)		(0.044)		(0.066)		(0.042)
P(Attend FD) in 2008		0.440***				0.378***	
		(0.018)				(0.018)	
Treated Middle School				0.322***		0.140***	
				(0.019)		(0.017)	
F (weak test)			571		287		298
Hasen Test: p-value							0.072
N	207,624	207,624	207,624	207,624	207,624	207,624	207,624
Panel B. Language test scores							
Full-day $\times$ Grade$_{g=12}$ $\times$ Cohort$_{t=t^*}$	0.195***		0.103*		0.250***		0.133***
	(0.012)		(0.042)		(0.062)		(0.039)
P(Attend FD) in 2008		0.440***				0.378***	
		(0.018)				(0.018)	
Treated Middle School				0.322***		0.140***	
				(0.019)		(0.017)	
F (weak test)			571		287		298
Hasen Test: p-value							0.021
N	207,615	207,615	207,615	207,615	207,615	207,615	207,615

Table D.4
Difference-in-Differences Model: Effects of full-day high schools on student test scores

	Math	Language
	(1)	(2)
Full-day × $\text{Grade}_{g=9}$	0.00977	0.0742
	(0.0540)	(0.0396)
Full-day × $\text{Grade}_{g=12}$	0.294***	0.314***
	(0.0637)	(0.0448)
N	33381	33381

Notes: This table reports results from a difference-in-differences model that estimates the effects of the full-day high school program on student gains from elementary to middle school (grade 9) and from middle to high school (grade 12). It includes a dummy of grade, school fixed effects, student age, and student gender as covariates. Standard errors are shown in parenthesis and they are clustered at the school-grade level. ${}^{*}p < 0.05$, ${}^{**}p < 0.01$, ${}^{***}p < 0.001$

Table D.5
Differential Attrition by Regular and Full-day Schools

	Regular	Full-Day		Test for equality	
		Pre Program	Post Program	Chi2	p-value
	(1)	(2)	(3)	(4)	(5)
Math Score (sd) - 9th grade	-0.054***	-0.047***	-0.061***	5.269	0.072
	(0.001)	(0.004)	(0.004)		
Age 10th grade (years)	0.042***	0.050***	0.078***	16.160	0.000
	(0.001)	(0.002)	(0.002)		
Female	-0.069***	-0.071***	-0.071***	0.052	0.974
	(0.002)	(0.007)	(0.007)		
Minority	-0.005	0.003	-0.010	1.596	0.450
	(0.003)	(0.008)	(0.008)		
Mother completed secondary	-0.009**	-0.018	-0.011	0.846	0.655
	(0.003)	(0.009)	(0.009)		
CCT eligible	-0.014***	-0.021*	-0.008	1.120	0.571
	(0.003)	(0.009)	(0.009)		
Less than 2 bathrooms	0.006	0.006	-0.001	0.392	0.822
	(0.003)	(0.010)	(0.010)		
Computer in the household	-0.038***	-0.034***	-0.033***	0.453	0.797
	(0.003)	(0.009)	(0.008)		
Less than 20 books at home	0.001	0.004	-0.008	0.974	0.615
	(0.003)	(0.009)	(0.009)		
Unconditional Means	0.49	0.43	0.41	-	-
N	161758	18308	17302	-	-

Notes: This table reports characteristics of students for whom we observe their pre-program test scores but not their end test scores. Each row is a different student characteristic reported by the student during the state exam. Columns 4 and 5 report a test for equality of coefficients shown in Columns 1-3.

Table D.6

Heterogeneous effects of full-day high school program on student language test scores by status of principal replacement at the full-day schools

	Principal replacement in the full-day school			
	No replacement	Before the program	At the year the program started	After the program started
	(1)	(2)	(3)	(4)
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.233^{***}	0.108^{***}	0.230^{***}	0.168^{***}
	(0.0195)	(0.0259)	(0.0190)	(0.0309)
F-test for equality (prob > F)	0.0166			
N	176,999	175,647	181,513	172,839

Notes: This table reports the estimates of heterogeneous effects of the full-day high school program on student language test scores. It divides the full-day high schools into four groups. Column 1 reports the effects of the program for full-day schools that did not change their principal from 2009 to 2015. Column 2 reports the effects for full-day schools that changed their principal before the program started. Column 3 reports the effects for full-day schools that changed their principal in the year the program started. Finally, Column 4 reports the effects for full-day schools that changed their principal one or two years after the program started. The row "F-test for equality" reports the p-value for a test of equality across columns 1-4. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table D.7

Heterogeneous effects of full-day high school program on student language test scores by teacher average characteristics of full-day schools

	Female		Elite college		New Teacher	
	Below median	Above median	Below median	Above median	Below median	Above median
	(1)	(2)	(3)	(4)	(5)	(6)
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.233***	0.157***	0.217***	0.171***	0.200***	0.192***
	(0.0183)	(0.0145)	(0.0165)	(0.0164)	(0.0165)	(0.0165)
F-test for equality (prob > F)		0.0007		0.0411		0.7139
N	186,430	187,646	187,118	186,958	186,157	187,919
	Teaching more than one cohort		Number of classrooms		Number of schools	
	Below median	Above median	Below median	Above median	Below median	Above median
	(7)	(8)	(9)	(10)	(11)	(12)
Full-day × $\text{Grade}_{g=12}$ × $\text{Cohort}_{t=t^*}$	0.219***	0.171***	0.167***	0.216***	0.163***	0.223***
	(0.0167)	(0.0160)	(0.0184)	(0.0150)	(0.0150)	(0.0174)
F-test for equality (prob > F)		0.0336		0.0341		0.0094
N	186,462	187,614	184,536	189,540	186,058	188,018

Notes: This table reports estimates of the effects of the full-day high school program by teacher average characteristics teaching at full-day high schools. For each average characteristic, we divided schools into two groups: below the median and above the median. The columns low and high make references to these groups, respectively. All columns control for $school \times Grade$, $school \times Cohort$, $Cohort \times Grade$, and student characteristics. For more details, see notes in Table 8. Values in parenthesis show standard errors clustered at the school-year-grade levels. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$