Do extended school day programs affect performance in college admission tests?*

Daniel Araujo*  
Guilherme Bayma†  
Carolina Melo‡  
Milena Mendonça*  
Luciano Sampaio§

Abstract
We assess the effects of a Brazilian extended high school day program on college admission test scores. The program entails an increase in Math/Science and Language class time, and the introduction of extra-class activity time. We exploit variation in program implementation timing to apply a difference-in-differences strategy and an event-study approach. Results show positive large effects on test scores and suggest that these effects are likely driven by increased class and extra-class activity hours. Moreover, the program affects not only scores in subjects for which it contemplates increased class time, but also in Humanities. Finally, we are able to show that increased time dedicated to academic-related extra-class activities works as a multiplier of program effects. In fact, these additional hours spent in complementary activities make the magnitude of program effects double from 0.101–0.148 to 0.182–0.316 standard deviation.

Keywords: extended high school day program, difference-in-differences, event-study, college admission test scores
JEL Codes: I21, I23, I26

1 Introduction
Developing countries are on the run to improve education and catch up with the developed ones. Extended School Day (ESD) programs have emerged as a
solution to improve post-primary education in many Latin American countries. However, evidence from different countries shows no consensus with regards to the effectiveness of these programs in improving test scores (Alfaro, Evans, & Holland, 2015; Almeida, Bresolin, Borges, Mendes, & Menezes-Filho, 2016; Hincapie, 2016). Furthermore, since ESD programs usually encompass a package of school improvements that go beyond increased instruction time, most researchers have not been successful in showing which program components drive the effects they have found.

In this paper, we use a difference-in-differences (DD) strategy and an event-study approach to evaluate the effects of the ESD program of the state of Pernambuco, Brazil, on national college admission test (ENEM\textsuperscript{1}) scores. In practice, the ESD program of Pernambuco has consisted in the transitioning of public high school education from regular-time schooling to quasi-full-time or full-time schooling gradually, with new schools entering the program each year.\textsuperscript{2,3} The main difference in the curriculum of the quasi-full-time and the full-time versions of the program is that the latter includes 8.3 hours per week of complementary activities—workshops and studying/research activities—, which is not included in the curriculum of the former.

On one hand, by estimating the effects of the quasi-full-time and the full-time versions of program separately and then comparing the results, we are able to assess the effect of the additional 8.3 weekly hours of complementary activities. On the other hand, with a less trivial approach, we use an event-study strategy to take advantage of both the staggered implementation and another program feature to disentangle the effects of increased school time from other program components—such as compensation and improved infrastructure—: when a school enters the program in a given year, increased class and extra-class hours are only offered to first-year high school students, so college admission test-takers in the first two years of program implementation are those students who only enjoyed program components other than increased school time. Since, with the event-study approach, we are able to assess program effects in each year after implementation, we can identify the isolated effect of program components other

\textsuperscript{1} The national college admission test is officially called \textit{Exame Nacional do Ensino Medio} (ENEM) in Brazil.

\textsuperscript{2} In regular-time schools, students go to class five days a week in only one shift each day. In quasi-full-time schools, students attend school five days a week in one shift, and 3 days a week in two shifts. And, in full-time schools, students go to class in two shifts five days a week.

\textsuperscript{3} The increase in hours has been mainly used for Language (Portuguese), Math, and Science instruction. See Table 1 for a summary of the hours per school activity contemplated in regular schools and treated schools.
than increased time by looking at effects in the first two years into the program.

Unlike previous studies, which generally document effects on Math and Language performance, we find positive effects on scores in all five ENEM subjects: Language, Math, Sciences, Humanities, and Writing. Results show that the additional hours spent in complementary activities contemplated in the full-time version of the program make the magnitude of program effects double from 0.101–0.148 to 0.182–0.316 standard deviation. Furthermore, we provide evidence that increased class and extra-class hours are likely the main drivers of effects—rather than other program components—, because effects are only observed starting from the third year after the program implementation, and this result applies to both the quasi-full-time and the full-time versions of the program.

In general, our findings add to the literature that explores the effects of increased school time on test scores. Evidence in this literature presents mixed results. Jensen (2013) analyzes a ninth grade change in the number of class hours and documents a positive effect on Math but no significant effect on Literacy. Meyer and Van Klaveren (2013) use a randomized control trial in the Netherlands to show that a 3-month extended day program has no significant effect on Math or Language achievement. Battistin and Meroni (2016) apply a difference-in-differences approach to analyze a program that increases instruction time in low achieving schools, and find positive effects on Math but no effect on Literacy. Huebener, Kuger, and Marcus (2017) look at a large reform in Germany, which contemplates increased instruction time, and find that, on average, the reform improves student performance; however, treatment effects are small. Taylor, Oberle, Durlak, and Weissberg (2017) use a regression-discontinuity design to study the effect of an increase in the number of Math class hours. The authors provide evidence of positive effects on Math; however, two years later, only one-third of the initial treatment effect remains. Finally, Figlio, Holden, and Ozek (2018) study a Florida program also focused on low-performing schools and document significant positive effects of additional literacy instruction on student reading abilities. Evidence from Brazil alone also presents mixed results. Oliveira (2010) finds that extending the school day by one hour has a positive average effect on Math test scores. While Almeida et al. (2016), analyzing a different program, find that the ESD has no nation-wide average effect on Language test scores and has a negative nation-wide average effect on Math test scores.

More related to our analysis, Rosa, Bettinger, Carnoy, and Dantas (2020)
Do extended school day programs affect performance in college admission tests?

also evaluate the ESD program of the state of Pernambuco. Using data on Math and Language standardized test scores of students in the 9th and 12th grades, the authors use a triple difference approach to assess program effects. They compare students’ 12th grade test scores with their 9th grade test scores, treated versus untreated, and before versus after treatment implementation. Their analysis points that treated students have a significantly higher gain in Math and Language abilities over the high school years than untreated students do. Furthermore, by exploring the differences between the full-time version of the program and its quasi-full-time version, the authors show evidence that increased time dedicated to supportive extra-class activities makes the program more effective.

This work complements the analysis performed by Rosa et al. (2020) in a few ways. First, we assess program effects on college admission tests, rather than effects on a state-administered test—aimed at assessing whether students hold abilities and competencies that are considered essential for their educational stage (Secretaria de Educação de Pernambuco, 2020), but not necessarily essential for a successful performance in the national college admission test. Second, since we combine a difference-in-differences strategy with an event-study approach, we are able to estimate effects for different years after program implementation and isolate the effect of program components other than increased hours at school. Third, we use an identification strategy different from that used by Rosa et al. (2020) to find additional evidence that complementary extra-class activities actually magnify program effects. And, finally, by evaluating effects on performance in five school subjects—rather than just in Math and Language—, including one (Humanities) which is not contemplated with increased instruction time in the program, we find that the program has important positive externalities in student learning.

We emphasize that this work contributes to the economics of education literature for it provides new evidence that ESD programs improve students’ performance not only in the subjects for which increased instruction time is contemplated, but also in other disciplines. This new evidence suggests the existence of a potential complementarity among the disciplines in learning. This is an aspect of great importance, given that it could be the case that, for instance, students spend so much more time in activities related to Math, Sciences, and

---

4 Respectively, the year right before entry in high school and the year of exit of high school in Brazil.

5 Rosa et al. (2020) use data on Math and Language scores on a state-administered test, called SAEPE (Sistema de Avaliação do Estado de Pernambuco).
Language that they fail to focus on other subjects of the college admission test, such as History, Geography, Philosophy, and Sociology (Humanities). Failure to focus on other subjects could, in turn, negatively affect college admission. Finally, our paper brings additional evidence to the table regarding the discussion about the design of ESD programs. In line with the work of Rivkin and Schiman (2015), Lavy (2020), and Rosa et al. (2020), we address potential benefits derived from the interaction of increased instruction time with extra-class activities, improved classroom environment, and school resources.

It is worth noting that the policy implication of this study is not trivial. Although we find that ESD programs help students improve performance in college admission tests—which is an important result—, this effect may not necessarily result in lower inequality. If the program were scaled up to cover a region completely, then students trained in full-time schools would have no advantage with regards to performance in the national college admission test. In the case of Pernambuco, Brazil, if the ESD program were scaled up to all public high schools, it would help close the gap—or, at least, help prevent the gap from being larger than otherwise—between public school students and private school students.

The rest of the paper is organized as follows. section 2 provides background information on the program and its implementation, section 3 presents our data. Section 4 presents our empirical strategy. Section 5 presents our main results. Section 6 presents robustness checks. And section 7 concludes.

2 The Extended School Day Program of Pernambuco, Brazil

Pernambuco’s state government started its ESD high school program in 2008. The program has two distinct versions: full-time and quasi-full-time. That is, when entering the ESD program, a public school adopts either the program’s full-time version or its quasi-full-time version. On one hand, in the program’s full-time version, children go to high school every weekday in two shifts. On the other hand, in the program’s quasi-full-time version, children go to high school two weekdays in one shift and three weekdays in two shifts—in other words, in the program’s quasi-full-time version, the school day is extended only in three of the five weekdays.

In both full-time schools and quasi-full-time schools, there is an increase in Math/Science and Language instruction time by 52 and 19 percent respectively. While regular public high schools offer 8.6 hours of Math/Science classes per week
and 4.2 hours of Language classes per week, schools that are part of the program offer 13.1 hours of Math/Science classes per week and 5 hours of Language classes per week. Treated schools also offer increased hours of Physical Education and Arts’ classes—see Table 1 for more details. In addition to increments in class hours, schools that are part of the program also offer increased extra-class time. Extra-class activities include both “Diversified Activities” and “Complementary Activities”. On one hand, “Diversified Activities” involve activities related to Foreign Languages, Entrepreneurship Projects, and Human Rights. On the other hand, “Complementary Activities” consist of more academic-related ventures, including workshops and studying/research activities.

Quasi-full-time schools and full-time schools offer about 1.6 and 10 additional hours of extra-class activities, respectively, in comparison with regular schools.

Table 1. Summary of program design, in terms of hours per week dedicated to class and extra-class activities.

| Subject / Activity                              | Regular Schools | Treated Schools |
|------------------------------------------------|-----------------|-----------------|
|                                                 | Quasi-Full-Time | Full-Time       |
| Language (Portuguese)                          | 4.2 h/week      | 5.0 h/week      |
| Physical Education                             | 0.8 h/week      | 1.7 h/week      |
| Arts                                           | 0.6 h/week      | 1.1 h/week      |
| Mathematics                                    | 3.3 h/week      | 5.0 h/week      |
| Sciences (Chemistry, Physics, Biology)         | 5.3 h/week      | 8.1 h/week      |
| Humanities (History, Geography, Philosophy, Sociology) | 5.0 h/week | 5.0 h/week      |
| Diversified Activities (Foreign Languages, Entrepreneurship Projects, Human Rights)* | 1.7 h/week | 3.3 h/week |
| Complementary Activities (Workshops, Studying / Research Activities) | 0.0 h/week | 0.0 h/week |

Source: Estado de Pernambuco (2012).
Notes: The table presents the average number of hours per week dedicated to each subject/activity in regular schools and in treated schools (quasi-full-time and full-time schools). *Note that regular schools do not include the Entrepreneurship Project and Human Rights activities within their set of “Diversified Activities”.

6 Note that regular schools also offer “Diversified Activities”, but they only include activities related to Foreign Languages. Activities related to Entrepreneurship Projects and Human Rights are only included in the curriculum of quasi-full-time and full-time schools.
Regular schools offer only 1.7 hours of “Diversified Activities” per week—which contemplate only activities related to Foreign Languages—; quasi-full-time schools offer 3.3 hours of the full package of “Diversified Activities” per week; and full-time schools offer 3.3 hours of the full package of “Diversified Activities” plus 8.3 hours of the full package of “Complementary Activities” per week.

In addition to the increase in instruction time and extra-class time described above, the program also contemplates other school interventions. Principals who work at high schools that participate in the ESD program receive higher compensation than principals of regular public high schools, and they are more likely to be hired through a competitive process. Teachers of participating schools must be of exclusive dedication to their schools, and are paid a higher compensation than that paid to teachers of regular public high schools. Finally, the program also entails an improvement in school infrastructure, with the remodeling or construction of labs and study rooms (Estado de Pernambuco, 2012). Due to the characteristics of the program’s implementation, we can show that these side-improvements alone do not have effects on students’ college admission test scores.

Two features of the program implementation scheme are key for this study:

1. The program has been implemented gradually, with new public high schools becoming quasi-full-time or full-time each year. The program started with 47 new participating schools in 2008. With this number increasing year after year, the program had reached approximately 345 participating schools in 2018, of which 172 were quasi-full-time and 173 were full-time.

2. When a school first adopts the program in a given year, increased class and extra-class hours are only offered to first-year high school students. However, even though second- and third-year students are not contemplated with extended class and extra-class hours, they benefit from the other components of the program—improved teacher’s and principal’s compensation and improved infrastructure. In the second year of program implementation, participating schools start offering increased class and extra-class hours to both first- and second-year students; and, in the third year, schools offer increased class and extra-class hours to students in all three years of high school. Thus, it takes three years for the implementation of the program to be completed (SEP, 2019).7

---

7 Data on the number of participating schools and details on the program’s implementation...
Because students who take the national college admission test (ENEM) are usually the ones who are in their last year of high school, when a participating school is still finishing its first or second year of program implementation, test-takers are the ones who did not have the opportunity to enjoy increased class and extra-class hours. This aspect of the program implementation dynamics enables us to isolate the effect of increased class and extra-class activity time from the effect of other program components—the compensation and infrastructure components. Additionally, the fact that the program can be implemented either in its quasi-full-time or its full-time version allows us to assess the effect of an additional 8.3 hours of extra-class, academic-related, activities per week.

3 Data

We use data from two different sources: the National Institute of Educational Research (INEP) and Pernambuco’s State Secretariat for Education (SSE). From INEP’s website, we downloaded two publicly available datasets: the ENEM’s dataset (from 2009 up to 2018) and the School Census of 2008. From the ENEM’s dataset, we obtained data on college admission test (ENEM) scores, student characteristics, and household characteristics at the individual level per year. From the School Census of 2008, we obtained data on pretreatment characteristics of schools, used along with other pretreatment demographic characteristics of students to apply Hainmueller (2012)’s entropy balancing method as a robustness check exercise—see section 6 for more details. From the SSE, we obtained—in 2019, upon direct request—data at the school level per year, identifying whether each school was a full-time or a quasi full-time unit in each year. By combining these datasets, we built an individual-by-school-by-year panel from 2009 up to 2018.

In Table 2, we present summary statistics—means and standard deviations—of the main variables used in our main analyses. In Table A-2, we present summary statistics of variables that describe pretreatment school characteristics, used in the robustness check exercises that involve entropy balancing.

---

8 Available at http://inep.gov.br/microdados
9 Although the program started in 2008, quality data on college admission test (ENEM) scores is only available starting from 2009.
10 We dropped schools treated in years prior to 2009, so that we could have an adequate baseline group.
11 To make college admission test (ENEM) scores comparable to other educational outcomes used in the literature, we standardized them.
Table 2. Summary Statistics.

|                      | Regular Never Treated | Quasi-full-Time Program Treated >2 years before full treatment | Treated 1 to 2 years before full treatment | Treated post full treatment | Full-Time Program Treated >2 years before full treatment | Treated 1 to 2 years before full treatment | Treated post full treatment |
|----------------------|-----------------------|---------------------------------------------------------------|------------------------------------------|----------------------------|----------------------------------------------------------|------------------------------------------|----------------------------|
| 2-10                 | (1)                   | (2)                                                          | (3)                                      | (4)                        | (5)                                                      | (6)                                      | (7)                        |
| **Student Characteristics** |                       |                                                               |                                          |                            |                                                          |                                          |                            |
| Age (Years)          | 20.512                | 20.323                                                       | 19.599                                    | 18.509                      | 20.194                                                   | 19.657                                    | 17.969                     |
|                      | (4.142)               | (1.981)                                                      | (1.765)                                  | (1.172)                     | (2.138)                                                  | (1.749)                                  | (0.96)                     |
| Female               | 0.606                 | 0.641                                                        | 0.632                                    | 0.596                       | 0.623                                                    | 0.622                                    | 0.596                      |
|                      | (0.210)               | (0.083)                                                      | (0.082)                                  | (0.076)                     | (0.103)                                                  | (0.125)                                  | (0.067)                    |
| **Household Characteristics** |                       |                                                               |                                          |                            |                                                          |                                          |                            |
| Share of students with family income from 0 to 2 minimum wages | 0.875                | 0.81                                                         | 0.927                                    | 0.922                       | 0.853                                                   | 0.881                                    | 0.907                      |
|                      | (0.203)               | (0.213)                                                      | (0.106)                                  | (0.069)                     | (0.192)                                                  | (0.167)                                  | (0.063)                    |
| Share of mothers with 0 to 8 years of education    | 0.629                | 0.616                                                        | 0.687                                    | 0.608                       | 0.638                                                    | 0.625                                    | 0.567                      |
|                      | (0.261)               | (0.188)                                                      | (0.151)                                  | (0.15)                      | (0.172)                                                  | (0.175)                                  | (0.149)                    |
| Share of fathers with 0 to 8 years of education     | 0.684                | 0.693                                                        | 0.769                                    | 0.673                       | 0.697                                                    | 0.704                                    | 0.682                      |
|                      | (0.254)               | (0.209)                                                      | (0.155)                                  | (0.157)                     | (0.19)                                                   | (0.183)                                  | (0.149)                    |
| **Student Test Scores** |                       |                                                               |                                          |                            |                                                          |                                          |                            |
| Humanities Test Score | 483.763               | 466.502                                                      | 485.958                                  | 507.296                     | 478.795                                                  | 481.848                                   | 512.413                    |
|                      | (47.007)              | (30.432)                                                     | (30.08)                                  | (26.744)                    | (35.375)                                                  | (33.357)                                  | (28.694)                   |
| Sciences Test Score  | 446.579               | 437.771                                                      | 438.092                                  | 459.473                     | 442.238                                                  | 443.795                                   | 465.223                    |
|                      | (37.425)              | (19.202)                                                     | (15.924)                                 | (17.263)                    | (21.797)                                                  | (21.222)                                  | (20.296)                   |
| Language Test Score  | 472.676               | 457.38                                                       | 463.751                                  | 483.081                     | 464.23                                                   | 464.253                                   | 488.096                    |
|                      | (41.317)              | (23.411)                                                     | (20.257)                                 | (19.591)                    | (23.583)                                                  | (22.651)                                  | (20.794)                   |
| Math Test Score      | 459.687               | 449.993                                                      | 444.373                                  | 466.538                     | 446.024                                                   | 458.176                                   | 477.222                    |
|                      | (55.074)              | (21.48)                                                      | (24.245)                                 | (29.581)                    | (23.571)                                                  | (26.808)                                  | (31.233)                   |
| Writing Test Score   | 387.509               | 375.461                                                      | 365.505                                  | 450.82                      | 375.715                                                   | 381.389                                   | 480.72                     |
|                      | (133.071)             | (81.028)                                                     | (73.694)                                 | (63.062)                    | (90.033)                                                  | (87.77)                                   | (61.125)                   |

*Note*: The table presents means and standard deviations of variables used in this study.
4 Empirical Strategy

In order to assess the causal effect of the ESD program of Pernambuco, Brazil, on college admission test (ENEM) scores, we would have to, ideally, compare treated schools with themselves in the absence of treatment—that is, with the schools’ counterfactuals. In a potential outcomes framework—as in Rubin’s Model for Causal Inference (Holland, 1986)—, we use $T = \{1, 0\}$ to denote the treatment status and define $Y_{s,t}(T)$ as the potential outcome of school $s$, in year $t$, with treatment status $T$. Hence, $Y_{s,t}(1)$ represents the potential outcome of school $s$, in year $t$, in the case that school $s$ was treated, and $Y_{s,t}(0)$ represents the potential outcome of school $s$, in year $t$, in the case that school $s$ was not treated.

Given the potential outcomes, we are interested in two parameters: the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT). These parameters can be expressed as follows:

$$ATE = \mathbb{E}[Y_{s,t}(1)] - \mathbb{E}[Y_{s,t}(0)]$$

$$= \{\mathbb{E}[Y_{s,t}(1) \mid T_s = 1] \cdot \Pr[T_s = 1] + \mathbb{E}[Y_{s,t}(1) \mid T_s = 0] \cdot \Pr[T_s = 0]\}$$

$$- \{\mathbb{E}[Y_{s,t}(0) \mid T_s = 1] \cdot \Pr[T_s = 1] + \mathbb{E}[Y_{s,t}(0) \mid T_s = 0] \cdot \Pr[T_s = 0]\}$$

$$ATT = \mathbb{E}[Y_{s,t}(1) \mid T_s = 1] - \mathbb{E}[Y_{s,t}(0) \mid T_s = 1].$$

But note that we can only observe $\mathbb{E}[Y_{s,t}(1) \mid T_s = 1]$ and $\mathbb{E}[Y_{s,t}(0) \mid T_s = 0]$. These two observable terms alone allow us to obtain none of the parameters of interest. We could, naively, try to estimate the ATT by taking the difference between $\mathbb{E}[Y_{s,t}(1) \mid T_s = 1]$ and $\mathbb{E}[Y_{s,t}(0) \mid T_s = 0]$; however, because $\mathbb{E}[Y_{s,t}(0) \mid T_s = 1] \neq \mathbb{E}[Y_{s,t}(0) \mid T_s = 0]$, we would obtain biased results. In fact, we have that

$$\mathbb{E}[Y_{s,t}(1) \mid T_s = 1] - \mathbb{E}[Y_{s,t}(0) \mid T_s = 0]$$

$$= \left\{\mathbb{E}[Y_{s,t}(1) \mid T_s = 1] - \mathbb{E}[Y_{s,t}(0) \mid T_s = 1]\right\}$$

$$+ \left\{\mathbb{E}[Y_{s,t}(0) \mid T_s = 1] - \mathbb{E}[Y_{s,t}(0) \mid T_s = 0]\right\},$$

where $\left\{\mathbb{E}[Y_{s,t}(1) \mid T_s = 1] - \mathbb{E}[Y_{s,t}(0) \mid T_s = 1]\right\}$ is the actual ATT of interest and $\left\{\mathbb{E}[Y_{s,t}(0) \mid T_s = 1] - \mathbb{E}[Y_{s,t}(0) \mid T_s = 0]\right\}$ consists of selection bias.
selection bias may be so large (in absolute value) that it completely masks a positive treatment effect” (Angrist & Pischke, 2009).

With a naive empirical strategy, such as a simple OLS regression, through which we assess the average difference in outcomes between treated and untreated schools using observational data, we are likely to obtain biased results due to selection bias. Even if OLS is improved with the inclusion of controls—such as student and school characteristics—, one could always argue that not all variables that affect both the treatment status and the outcome of interest have been taken into account, compromising the internal validity of the results. Furthermore, we could also face the problem of over-controlling, since coefficients are also biased when a control is caused by the explanatory variable of interest.

To circumvent the Fundamental Problem of Causal Inference (Holland, 1986)—that is, the problem of not being able to observe the counterfactuals needed to perfectly identify the ATE or the ATT—we use an identification strategy based on the Difference-in-Differences (DD) approach. This strategy allows us to identify the ATT, given that a certain identification assumption holds.

We still use $T = \{1,0\}$ to denote the treatment status, and now we also use $t = \{1,0\}$ to denote the periods post- and pre-treatment, respectively. Then, the DD estimator is given by

$$
\beta_{\text{DD}} = \left\{ E[Y_{s,t} | T_s = 1, t = 1] - E[Y_{s,t} | T_s = 1, t = 0] \right\} - \left\{ E[Y_{s,t} | T_s = 0, t = 1] - E[Y_{s,t} | T_s = 0, t = 0] \right\}.
$$

(4)

In order for $\beta_{\text{DD}}$ to actually represent the ATT of interest, the following identification assumption must hold:

$$
E[Y_{s,t}(0) | T_s = 1, t = 1] - E[Y_{s,t}(0) | T_s = 1, t = 0] \\
= E[Y_{s,t}(0)T_s = 0, t = 1] - E[Y_{s,t}(0)T_s = 0, t = 0].
$$

(5)

That is, when applying the DD strategy, we assume that the pre- and post-treatment difference in outcomes for the control group is equal to the pre- and post-treatment difference in outcomes for the treatment group, had it not been treated. In other words, we assume that the outcome variable trends for the control and treatment groups would be parallel, had the program not been implemented.

The DD strategy is powerful for it allows us to control for all possible
Do extended school day programs affect performance in college admission tests?

Time-invariant characteristics of schools and all factors that change over time equally for treated and untreated schools, even when these characteristics and/or factors are unobserved. However, this strategy does not allow us to control for time-variant shocks that affect treated and untreated schools differently. Therefore, issues related to student migration and compositional changes in the profile of students in treated and untreated schools over time, for instance, can represent threats to the internal validity of this study. We address some of these threats in section 6, with robustness check exercises.

Now, we move forward to explain how our DD strategy is implemented. But, first, it is worth noting that, because treated schools enter the program in different time periods, the natural way for us to implement a DD strategy is through a two-way fixed effects model. Since students usually take the national college admission test (ENEM) in their last year of high school, the first completely treated group of students in a participating school takes the test only two periods after the school entered the program. Hence, in a given year $t$, untreated public schools and public schools that are in the first two years of program implementation are considered controls, while public schools that entered the program in year $t - 2$ or earlier are considered treated. This idea is generalized in a DD regression framework expressed by the following specification:

$$\text{TestScore}_{i,s,t} = \theta_s + \lambda_t + \beta \text{Treatment}_{s,t} + \epsilon_{i,s,t},$$

(6)

where $\text{TestScore}_{i,s,t}$ represents the test score for the student $i$, who studied at school $s$, and took the test in year $t$; $\theta_s$ and $\lambda_t$ are school and time fixed effects respectively; $\epsilon_{i,s,t}$ is the random error, clustered at the school level to account for potential serial correlation; $\text{Treatment}_{s,t}$ is a dummy variable that takes the value of 1 if the school $s$ was fully treated in year $t$ (that is, if the school $s$ had reached full program implementation in year $t$) and the value of zero otherwise; and, finally, $\beta$ represents the average treatment effect on the treated—which refers to either the full-time or the quasi full-time model.

Along with estimations of the above DD regression specification, for the purpose of comparison, we also estimate equation (6) with student demographic characteristics as controls and simple OLS regressions, with and without student demographic characteristics as controls. On one hand, the inclusion of student demographic characteristics as controls in our main DD specification allows us to verify whether the omission of student characteristics that (potentially) change over time can result in biased estimates of the parameter of interest. On the other
hand, the estimation of naive OLS regressions allows us to show how completely ignoring the Fundamental Problem of Causal Inference (Holland, 1986) leads us to the under- or over-estimation of the effects of interest. As one will note, results of our main DD specification change very little with the inclusion of controls. Naive OLS estimations, however, seem to produce coefficients that are negatively biased, even with the inclusion of demographic control variables. Results are discussed in greater depth in section 5.

In light of the recent literature on the decomposition of two-way fixed effects estimates with variation in treatment timing (Goodman-Bacon, 2018), we recognize that regression equation (6) only recovers the ATT of interest when treatment effects are homogeneous across units. Otherwise, units that are treated in the middle of the considered evaluation window may receive more weight. Furthermore, if treatment effects change over time within unit, our two-way fixed effects estimates could be biased. Our approach to addressing these problems are two-fold: i) we run robustness check exercises, using Goodman-Bacon, Goldring, and Nichols (2019)’s STATA command, to verify the sources of variation and the heterogeneity of treatment effects—see section 6 for more details—; and ii) we use an event-study specification as well, allowing treatment effects to change over time within unit, without compromising the validity of treatment effect estimates.

Then, with the objective of both addressing the potential bias of two-way fixed effects estimates and assessing treatment effects in different periods after program implementation—including the first two years of program adoption, when college admission test-takers are those who only enjoyed program components other than increased class and extra-class time—, we also use the event-study specification (Autor, 2003) below:

$$TestScore_{i,s,t} = \theta_s + \lambda_t + \sum_{j=-5}^{5} \beta_j \cdot \mathbf{1}(T_{st}^* - t = j) + \epsilon_{i,s,t},$$

where the indicator $\mathbf{1}(T_{st}^* - t = j)$ is a decomposition of the traditional treatment variable and measures time relative to the treatment year. We set the coefficient $\beta$ equal to zero for periods 6 years or more before the intervention, so that these years can be used as the baseline period. This choice of baseline is preferred over (the traditionally used) period $-1$ to allow for assessing the effect of program components other than increased class and extra-class activity hours. The parameters $\beta_{-2}$ and $\beta_{-1}$ represent the effect of the program in the first and second year of program implementation, respectively. These are the parameters
that allow us to identify the effect of improved compensation and improved infrastructure alone on college admission test (ENEM) scores of students that enjoyed these program components for one or two years.

5 Results

In tables 3 and 4, we present results of naive OLS estimations without and with demographic controls, and results of our main DD specification—equation (6)—without and with demographic controls, for the full-time and quasi-full-time versions of the program respectively. As one can note, simple OLS regressions produce coefficients that are negatively biased, even after controlling for demographic characteristics of students. Estimates based on equation (6), however, show that program effects are positive and statistically significant for all five college admission test (ENEM) subjects. On one hand, schools whose students were exposed to the full-time version of the program experienced an average increase of 0.182–0.316 standard deviation (s.d.) in college admission test (ENEM) scores. Schools whose students were exposed to the quasi-full-time version of the program, on the other hand, experienced an average increase of 0.101–0.148 s.d. in test scores. The larger effects of the full-time version of the program suggest that the hours of “Complementary Activities” work as a multiplier of program effects. Effects are very robust to the inclusion of demographic controls.

In Figure 1, we present event-study estimates of program effects on college admission test (ENEM) scores. Recall that treated schools entered the program in period $-2$ and the first period in which a fully treated cohort of these schools took the test was period 0. So, students from treated schools who took the test in periods $-2$ or $-1$ benefited only from program components other than increased class and extra-class activity hours. And students from treated schools who took the test in period 0, or in periods after 0, benefited from the whole package of school improvements contemplated by the program.

Results presented in Figure 1 show that program effects are mostly non-significant in periods $-2$ and $-1$—and, when they are, they are negative—but they are significant, positive, and large starting from period 0, for all five college admission test (ENEM) subjects. Furthermore, Figure 1 shows that the effects for full-time schools are about twice as large as the effects for quasi-full-time schools (consistent with results presented in tables 3 and 4) and, for both versions of the program, it does not seem that the magnitude of effects change significantly over time. Therefore, we can say that: i) improved compensation and infrastructure alone do not affect college admission (ENEM) test scores, so increased class
Table 3. Estimated Effects of the Extended School Day Program on college Admission Test (ENEM) Scores: Full-Time Program.

|          | Math          | Sciences        | Humanities        |
|----------|---------------|-----------------|-------------------|
|          | OLS (1)       | OLS (2)         | OLS (5)           |
|          | DD (3)        | DD (4)          | DD (7)            |
| Treatement| 0.0983***     | 0.206***        | 0.0888***         |
|          | (0.0351)      | (0.0205)        | (0.0332)          |
|          | 0.0515        | 0.123***        | 0.219***          |
|          | (0.0348)      | (0.0200)        | (0.0331)          |
|          | 0.206***      | 0.192***        | 0.208***          |
|          | (0.0205)      | (0.0200)        | (0.0238)          |
|          | 0.0936***     | 0.0936***       | 0.0936***         |
|          | (0.0319)      | (0.0320)        | (0.0217)          |
| Observations | 311,938      | 316,477         | 320,451           |
| R-squared | 0.001         | 0.002           | 0.001             |
| School/Year FE | No          | No              | No               |
| Controls  | No            | Yes             | No               |
| Language  | OLS (13)      | OLS (14)        | OLS (17)          |
|          | DD (15)       | DD (16)         | DD (19)           |
| Treatement| 0.107***      | 0.199***        | 0.186***          |
|          | (0.0332)      | (0.0249)        | (0.0241)          |
|          | 0.0609*       | 0.186***        | 0.189***          |
|          | (0.0333)      | (0.0249)        | (0.0236)          |
|          | 0.186***      | 0.199***        | 0.189***          |
|          | (0.0249)      | (0.0249)        | (0.0234)          |
|          | 0.297***      | 0.316***        | 0.258***          |
|          | (0.0222)      | (0.0234)        | (0.0234)          |
| Observations | 315,912      | 315,912         | 355,063           |
| R-squared | 0.002         | 0.169           | 0.183             |
| School/Year FE | No          | No              | No               |
| Controls  | No            | Yes             | No               |

Notes: Models presented are either naïve OLS estimates or DD estimates of equation (6). Even columns include controls for gender, age and a non-white dummy. Standard errors are clustered at the school level.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
Table 4. Estimated Effects of the Extended School Day Program on College Admission Test (ENEM) Scores: Quasi-Full-Time Program.

|                      | Math Sciences | Sciences | Humanities |
|----------------------|---------------|----------|------------|
|                      | OLS (1)       | OLS (2)  | DD (3)     | DD (4)     | OLS (5) | OLS (6)  | DD (7) | DD (8) |
| Treatment            | −0.0463       | −0.0788  | 0.0954***  | 0.0852***  | −0.0308 | −0.0535**| 0.103***| 0.0958***|
|                      | (0.0300)      | (0.0292) | (0.0140)   | (0.0135)   | (0.0275)| (0.0271) | (0.0125)| (0.0126)|
| Observations         | 326,388       | 317,911  | 326,388    | 317,911    | 331,186 | 322,540  | 331,186| 322,540|
| R-squared            | 0.000         | 0.062    | 0.204      | 0.243      | 0.000   | 0.032    | 0.183  | 0.199  |
| School/Year FE       | No            | No       | Yes        | Yes        | No      | No       | Yes    | Yes    |
| Controls             | No            | Yes      | No         | Yes        | No      | Yes      | No     | Yes    |

|                      | Language      | Writing   |
|----------------------|---------------|-----------|
|                      | OLS (13)      | OLS (14)  | DD (15)    | DD (16)    |
| Treatment            | −0.0135       | −0.0453*  | 0.109***   | 0.0985***  |
|                      | (0.0277)      | (0.0271)  | (0.0127)   | (0.0125)   |
| Observations         | 330,561       | 321,958   | 330,561    | 321,958    |
| R-squared            | 0.000         | 0.028     | 0.162      | 0.176      |
| School/Year FE       | No            | No        | Yes        | Yes        |
| Controls             | No            | Yes       | No         | Yes        |

Notes: Models presented are either naive OLS estimates or DD estimates of equation (6). Even columns include controls for gender, age and a non-white dummy. Standard errors are clustered at the school level.

***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.
Figure 1. Event-Study Estimates of the Effects of the Extended School Day Program on College Admission Test (ENEM) Scores.

Notes: This figure shows plots of event-study estimates—based on specification (7)—and corresponding 95-percent confidence intervals. Standard errors are clustered at the school level.
and extra-class activity hours are likely the components that drive program effects; ii) the additional 8.3 weekly hours of extra-class time—dedicated to workshops and studying/research activities—offered in the full-time version of the program are, in fact, able to almost double positive effects on test scores; and iii) results of our two-way fixed effects model do not seem to be harshly biased, since treatment effects do not change much over time after program implementation.

In a few cases—see the case of full-time schools in Sciences and quasi-full-time schools in Math—, Figure 1 reveals that the program has a small negative effect on college admission test (ENEM) scores in its first two years (periods $-2$ and $-1$). These small significant effects during the transition from a regular to a full-time or quasi-full-time school raise a concern regarding the results obtained from the estimation of our two-way fixed effects model—equation (6). Since we consider as controls public schools that are in the first two years of program implementation, it could be that these negative effects in periods $-2$ and $-1$ are magnifying the coefficients presented in tables 3 and 4. In order to address this problem, we run our two-way fixed effects model excluding periods $-2$ and $-1$ from the analysis—see section 6 for more details—and compare the results with those presented in tables 3 and 4. It turns out that, with the exclusion of periods $-2$ and $-1$, the coefficients of interest remain very stable, indicating that these effects that we observe for the first two years of program implementation are not strong enough to magnify average positive effects found for periods after full program implementation.

Figure 1 also shows that, in periods that precede program implementation—period $-3$ and earlier—, the treatment and control groups behave similarly in the vast majority of the cases, with non-significant test score differences between the two. These results suggest that trends in outcome variables were parallel before treatment implementation, strengthening our confidence in the assumption that they would have remained parallel had the ESD program not been adopted.

### 6 Robustness Checks

One potential threat to the internal validity of our study is student migration between schools,\textsuperscript{12} which could change the student profile of treated and untreated

\textsuperscript{12} Rosa (2019) uses a quasi-experimental approach to evaluate the effect of Pernambuco’s ESD program on enrollment in private high schools. The author shows that the program reduces enrollment in private schools and that the effect is likely driven by migration of students from private high schools to public full-day schools.
schools over time. To address this issue, we estimate equation (7) using the educational level of students’ parents and the level of family income as outcome variables.\textsuperscript{13} Results, presented in Figure A-1, show that the treatment has small negative effects on the proportion of mothers with up to 8 years of education and on the proportion of students with family income of up to 2 minimum wages. However, these effects only show up as significant four years after full program implementation. These results are consistent with Rosa (2019), who finds that the program has a small effect on student migration from private to public high schools in the first years after program implementation, but the effect only gets larger in magnitude as years pass. We, therefore, understand that the influence of such compositional changes is smaller closer to the treatment full implementation threshold (period 0), so we encourage the reader to focus on program effects in periods 0 and 1. Since we find large program effects on college admission test (ENEM) scores right after program implementation—see Figure 1—, when migration is not yet intense enough to change the profile of students in treated public schools, we can conclude that our results are mostly not driven by this compositional endogeneity and our main conclusions should hold.

By looking at Figure A-1, one could also argue that the proportion of mothers with up to 8 years of education and the proportion of fathers with up to 8 years of education were greater for treated schools before the adoption of the program. However, we do not attribute much importance to these pretreatment effects, because they are very small and mostly statistically insignificant.

Now, to check whether estimates are driven by a particular year, we rerun the baseline specification of equation (6) for the full-time and quasi-full-time versions of the program repeatedly, but each time excluding a specific group of observations. The results—presented in figures A-2 and A-3—, show that treatment effects are remarkably stable to exclusions of different years, suggesting that our main findings are not being distorted by any specific time period.

Furthermore, to ensure that our results cannot be easily reproducible at random, we perform permutation tests by randomly assigning each of the treated schools a false campaign year, and then estimating the model using the implied placebo treatment probability. This procedure is repeated 1,000 times. Figures A-4 and A-5 present the empirical cumulative distribution functions (CDFs) of placebo coefficients for all five test subjects. For all test subjects, the true coefficients fall above the 99th percentile of the distribution of placebo

\textsuperscript{13} In this case, we estimate the effects for the two versions of the program jointly.
coefficients, suggesting that our results are very unlikely to be obtained due to the randomness of the data.

As mentioned in section 5, although they are small, some significant program effects found for the first two years of program implementation (periods $-2$ and $-1$ in our event-study analysis) represent a concern to us regarding the results obtained from our two-way fixed effects model. To circumvent this issue, we rerun the estimations of equation (6), but now excluding the two periods that precede full program implementation. Results are presented in Table A-1. For each college admission test (ENEM) subject, we show program effect estimates with the inclusion of periods $-2$ and $-1$—odd columns—and without the inclusion of these periods—even columns. As one can note, estimates of treatment effects remain very robust to the exclusion of these years that precede full program implementation, suggesting that the small effects found in the event-study analysis for these periods are not capable of magnifying the average program effects obtained from our two-way fixed effects model.

Another possible concern regarding our baseline results relate to a recent set of methodological studies, highlighting some challenges in the estimation of two-way fixed effects models in the presence of differences in the treatment timing between unit groups (i.e., earlier treated, later treated, and never treated groups) (Goodman-Bacon, 2018; Goodman-Bacon et al., 2019). In this environment, the DD estimator represents a weighted average of all $2 \times 2$ estimators in the data, and thus all groups can be control units in some of the weighted estimates, and not just the group never treated. If we do not have a group of units never treated, or if this group is small, estimations can be driven by the weight of non-optimal comparisons (i.e., earlier group treated vs. later group control, or later group treated vs. earlier group control), generating biased estimates.

To check if this problem occurs in our estimations, we perform robustness check exercises based on Goodman-Bacon (2018) and Goodman-Bacon et al. (2019) to verify the impact of the presence of treatment timing on the weights of our two-way fixed effects model, represented by equation (6). Results for Goodman-Bacon (2018)’s two-way fixed effects decomposition for the full-time and the quasi-full-time programs are presented in figures A-6 and A-7, respectively. These figures present the betas for comparisons between all groups in the data and their respective weights in the main estimation. Results highlight that, due to the presence of a large group never treated in our sample, the largest fraction of the regression weights are used for comparisons between the groups of treated and never treated units. Therefore, it is very unlikely that our baseline results
are biased due to a large weight being assigned to comparisons of groups other than treated vs. never treated schools.

Finally, to improve our confidence in the results presented in this study, we also apply an additional empirical strategy to assess the effect of the ESD program of Pernambuco, Brazil. We combine our DD and event-study strategies with entropy balancing, “a data preprocessing method to achieve covariate balance in observational studies with binary treatment” (Hainmueller, 2012). As in other commonly used methods, such as matching and propensity score, the idea of applying entropy balancing is to make it more likely that treatment status is as good as random, given the pretreatment characteristics used in covariate balancing. Results obtained from the combination of the DD strategy with entropy balancing are presented in Table A-3, and results obtained from the combination of the event-study strategy with entropy balancing are presented in Figure A-8. The application of entropy balancing to achieve pretreatment covariate balance barely changes our results. In fact, results show that treatment effects remain practically unchanged and our main conclusions remain valid.

7 Concluding Remarks

In this paper, we evaluate an Extended School Day program implemented in Pernambuco, Brazil. Aside from providing evidence of program effects on college admission test scores, our contribution to the existing literature is threefold.

First, we show that increased class and extra-class activity time are likely the drivers of effects in extended day programs, rather than other program components, such as principals’ and teachers’ compensation and school infrastructure improvements. Second, we show that extra-class time—mostly dedicated to academic-related activities—can serve as a multiplier of the effects driven by the increased class time. In fact, these additional hours spent in complementary activities make the magnitude of program effects double from 0.101–0.148 to 0.182–0.316 standard deviation. And third, we reveal that offering more hours of instruction only in Math/Sciences and Language not only improves test scores in these subjects, but also generates positive externalities in other disciplines, such as History, Geography, Philosophy, and Sociology (Humanities).

To conclude, it is worth noting that the policy implication of this study is not obvious, since the extending of the school day by all schools in a territory would result in no advantage in college admission tests for treated students. Therefore, if the aim of the government were to reduce inequality, an Extended School Day
Do extended school day programs affect performance in college admission tests?

Policy should be focused on those schools that serve the most disadvantaged students.

References

Alfaro, P., Evans, D. K., & Holland, P. (2015, June). Extending the school day in Latin America and the Caribbean (Working Paper No. WPS7309). Washington, DC: World Bank. http://hdl.handle.net/10986/22183

Almeida, R., Bresolin, A., Borges, B., Mendes, K., & Menezes-Filho, N. (2016). Assessing the impacts of Mais Educação on educational outcomes: Evidence between 2007 and 2011 (Working Paper No. WPS7644). Washington, DC: World Bank. http://hdl.handle.net/10986/24218

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist’s companion. Princeton University Press.

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. Journal of Labor Economics, 21(1), 1–42. http://dx.doi.org/10.1086/344122

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

Estado de Pernambuco – Poder Executivo. (2012, February 28). Instrução normativa nº 01/2012: Fixa normas para a reorganização das Matrizes Curriculares da Educação Básica no âmbito das Escolas da Rede Estadual de Ensino de Pernambuco, para o ano letivo de 2012. Diário Oficial do Estado de Pernambuco, 38, 8–11. http://200.238.105.211/cadernos/2012/20120228/1-PoderExecutivo/PoderExecutivo(20120228).pdf

Secretaria de Educação de Pernambuco. (2020). SAEPE: Secretaria de Educação e Esportes divulga resultados de 2019. Retrieved 14-09-2020, from http://www.educacao.pe.gov.br/portal/?pag=1&cat=37&art=5570

Figlio, D., Holden, K. L., & Ozek, U. (2018). Do students benefit from longer school days? Regression discontinuity evidence from Florida’s additional hour of literacy instruction. Economics of Education Review, 67, 171–183. http://dx.doi.org/10.1016/j.econedurev.2018.06.003

Goodman-Bacon, A. (2018, September). Difference-in-differences with variation in treatment timing (Tech. Rep. No. 25018). National Bureau of Economic Research (NBER). http://dx.doi.org/10.3386/w25018

Goodman-Bacon, A., Goldring, T., & Nichols, A. (2019). BACONDE-COMP: Stata module to perform a Bacon decomposition of difference-in-differences estimation. https://ideas.repec.org/c/boc/bocode/s458676.html (Statistical Software Components S458676, Boston College Department of Economics)
Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20*(1), 25–46.  
http://dx.doi.org/10.1093/pan/mpr025

Hincapie, D. (2016, March).  *Do longer school days improve student achievement? Evidence from Colombia* (Working Paper No. IDB-WP-679). Washington, D.C.: Inter-American Development Bank (IDB).  
https://publications.iadb.org/en/do-longer-school-days-improve-student-achievement-evidence-colombia

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.  
http://dx.doi.org/10.2307/2289064

Huebener, M., Kuger, S., & Marcus, J. (2017). Increased instruction hours and the widening gap in student performance. *Labour Economics, 47*, 15–34.  
http://dx.doi.org/10.1016/j.labeco.2017.04.007

Jensen, V. M. (2013). Working longer makes students stronger? The effects of ninth grade classroom hours on ninth grade student performance. *Educational Research, 55*(2), 180–194.  
http://dx.doi.org/10.1080/00131881.2013.801244

Lavy, V. (2020). Expanding school resources and increasing time on task: Effects on students’ academic and noncognitive outcomes. *Journal of the European Economic Association, 18*(1), 232–265.  
http://dx.doi.org/10.1093/jeea/jvy054

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

Oliveira, J. M. (2010). Custo-efetividade de políticas de redução do tamanho da classe e de ampliação da jornada escolar: Uma aplicação de estimadores de matching. *Revista do BNDES, 33*, 281–306.  
http://web.bnides.gov.br/bib/jspui/handle/1408/3345

Rivkin, S. G., & Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal, 125*(588), F425–F448.  
http://dx.doi.org/10.1111/ecoj.12315

Rosa, L. (2019, November 9). The unintended consequences of public school subsidies on educational markets. In *The 24th Annual LACEA Meeting 2019*, Puebla, Mexico.  
https://sistemas.colmex.mx/Reportes/LACEALAMES/LACEA-LAMES2019_paper_331.pdf

Rosa, L., Bettinger, E., Carnoy, M., & Dantas, P. (2020). *The effects of public high school subsidies on student test scores: The case of a full-day high school program in Pernambuco, Brazil*.  
https://www.sonhogrande.org/storage/the-effects-of-public-high-school-subsidies-on-student-test
Do extended school day programs affect performance in college admission tests?

SEP – Secretaria de Educação de Pernambuco. (2019). *Plano Estadual de Educação, Pernambuco, 2015–2025*. https://legis.alepe.pe.gov.br/texto.aspx?id=10998&tipo=TEXTORIGINAL (LEI Nº 15.533, DE 23 DE JUNHO DE 2015)

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156–1171. http://dx.doi.org/10.1111/cdev.12864
Appendix A

Figure A-1. Event-Study Estimates of the Effects of the Extended School Day Program on the Profile of Students.

(a) Proportion of mothers with 0 to 8 years of education.

(b) Proportion of fathers with 0 to 8 years of education.

(c) Proportion of students with family income from 0 to 2 minimum wages.

Notes: This figure shows plots of event-study estimates—based on specification 7—and corresponding 95-percent confidence intervals. Standard errors are clustered at the school level. Fixed effects for school, year, and gender included.
Figure A-2. Robustness to Sample Restrictions: Dropping Cohorts (Full-Time Program).

Notes: This figure shows plots of point estimates for full-time treatment effects—based on specification (6)—and corresponding 95-percent confidence intervals, as different subsets of observations are excluded from the sample.
Figure A-3. Robustness to Sample Restrictions: Dropping Cohorts Quasi-Full-Time Program.

Notes: This figure shows plots of point estimates for quasi full-time treatment effects—based on specification (6)—and corresponding 95-percent confidence intervals, as different subsets of observations are excluded from the sample.
Figure A-4. Robustness to Sample Restrictions: Permutation Tests (Full-Time Program).

Notes: These figures show plots of the empirical CDF of placebo coefficients. We randomly assign each of the Full-time schools across cities a treatment year, and estimate model 1 using the implied placebo treatment. This procedure is repeated 1,000 times. The share of the 1000 absolute placebo coefficients that are larger than the absolute actual coefficient is the p-value for the hypothesis that $\beta = 0$. 
Figure A-5. Robustness to Sample Restrictions: Permutation Tests (Quasi-Full-Time Program).

*Notes:* These figures show plots of the empirical CDF of placebo coefficients. We randomly assign each of the Quasi Full-time schools across cities a treatment year, and estimate model 1 using the implied placebo treatment. This procedure is repeated 1000 times. The share of the 1000 absolute placebo coefficients that are larger than the absolute actual coefficient is the p-value for the hypothesis that $\beta = 0$. 
Table A-1. Estimated Effects of the Extended School Day Program on College Admission Test (ENEM) Scores – Dropping Years Preceding Full Program Implementation.

|                   | Math   | Sciences | Humanities | Language | Writing |
|-------------------|--------|----------|------------|----------|---------|
|                   | (1)    | (2)      | (3)        | (4)      | (5)     |
| **Full-Time Program** |        |          |            |          |         |
| Treatment         | 0.206*** | 0.212*** | 0.219***   | 0.230*** | 0.182*** |
|                   | (0.0205) | (0.0214) | (0.0238)   | (0.0245) | (0.0217) |
| Observations      | 311,938 | 300,644  | 316,477    | 304,991  | 320,451  |
| R-squared         | 0.211   | 0.216    | 0.190      | 0.194    | 0.173    |
| School/Year FE    | Yes     | Yes      | Yes        | Yes      | Yes      |
| Periods –1 and –2 excluded | No  | Yes      | No         | Yes      | No       |

|                   | (11)   | (12)     | (13)       | (14)     | (15)    |
| **Quasi Full-Time Program** |       |          |            |          |         |
| Treatment         | 0.0954*** | 0.0900*** | 0.103***   | 0.102*** | 0.101*** |
|                   | (0.0140) | (0.0135) | (0.0125)   | (0.0129) | (0.0130) |
| Observations      | 326,388 | 312,537  | 331,186    | 317,031  | 335,359  |
| R-squared         | 0.204   | 0.209    | 0.183      | 0.188    | 0.166    |
| School/Year FE    | Yes     | Yes      | Yes        | Yes      | Yes      |
| Periods –1 and –2 excluded | No  | Yes      | No         | Yes      | No       |

Notes: Models presented are least-squares estimates of equation (6). For each discipline, the first two columns represent OLS estimates and the last two difference-in-differences estimates with fixed effects for school and time. Even columns include also controls for gender, age and a non-white dummy. Standard errors are clustered at the school level.

***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.
Figure A-6. Two-Way Fixed Effects Weights Decomposition: Full-Time Program.

Notes: This figure plots results from Goodman-Bacon (2018) decomposition of our baseline two-way fixed effects estimates. More specifically, graphs show point estimates of the two-way fixed effects model in the y-axis, considering the comparison of different groups of treated and untreated units—earlier group treatment vs. later group control, later group treatment vs. earlier group control, and treatment vs. never treated—and the weight attributed to each of these group comparisons in the x-axis. The solid horizontal line represents our baseline estimate, considering all groups of treated and untreated units, with their respective weights, for matters of comparison. The 2 \times 2DD estimates and weights were obtained using STATA package bacondecomp.
Do extended school day programs affect performance in college admission tests?

Figure A-7. Two-Way Fixed Effects Weights Decomposition: Quasi-Full-Time Program.

Notes: This figure plots results from Goodman-Bacon (2018) decomposition of our baseline two-way fixed effects estimates. More specifically, graphs show point estimates of the two-way fixed effects model in the y-axis, considering the comparison of different groups of treated and untreated units—earlier group treatment vs. later group control, later group treatment vs. earlier group control, and treatment vs. never treated—and the weight attributed to each of these group comparisons in the x-axis. The solid horizontal line represents our baseline estimate, considering all groups of treated and untreated units, with their respective weights, for matters of comparison. The $2 \times 2$DD estimates and weights were obtained using STATA package bacondecomp.
### Table A-2. Summary Statistics: Pretreatment School Characteristics.

|                                      | Treated Schools | Control Schools |                  |                  |
|--------------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                      | Mean            | Variance        | Mean            | Variance        |
| **Without weighting**                |                 |                 |                 |                 |
| Presence of Computer Lab             | 0.900           | 0.090           | 0.787           | 0.168           |
| Presence of Sciences Lab             | 0.406           | 0.242           | 0.301           | 0.211           |
| Presence of Kitchen                  | 0.979           | 0.021           | 0.876           | 0.108           |
| Presence of Library                  | 0.929           | 0.066           | 0.909           | 0.083           |
| Number of Classrooms                 | 13.210          | 24.2            | 14.61           | 114.7           |
| Number of Employees                  | 55.320          | 366.5           | 58.62           | 2988            |
| Share of mothers with 0-8 yrs of education | 0.437   | 0.017           | 0.356           | 0.046           |
| Share of fathers with 0-8 yrs of education | 0.459  | 0.021           | 0.369           | 0.042           |
| Share of students with family income of 0-2 min. wages | 0.534 | 0.019           | 0.454           | 0.045           |
| Number of Students doing ENEM        | 100.5           | 6840            | 62.54           | 4666            |
| **With weighting**                   |                 |                 |                 |                 |
| Presence of Computer Lab             | 0.900           | 0.090           | 0.900           | 0.090           |
| Presence of Sciences Lab             | 0.406           | 0.242           | 0.406           | 0.241           |
| Presence of Kitchen                  | 0.979           | 0.021           | 0.979           | 0.021           |
| Presence of Library                  | 0.929           | 0.066           | 0.929           | 0.066           |
| Number of Classrooms                 | 13.21           | 24.2            | 13.21           | 24.2            |
| Number of Employees                  | 55.32           | 366.5           | 55.32           | 366.8           |
| Share of mothers with 0-8 yrs of education | 0.437   | 0.017           | 0.437           | 0.017           |
| Share of fathers with 0-8 yrs of education | 0.459  | 0.021           | 0.459           | 0.021           |
| Share of students with family income of 0-2 min. wages | 0.534 | 0.019           | 0.534           | 0.019           |
| Number of Students doing ENEM        | 100.5           | 6840            | 100.5           | 6840            |

Notes: The table presents summary statistics of variables that describe pretreatment school characteristics, used as covariates in the application of Hainmueller (2012)’s entropy balancing method, described in section 6. The weighting presented here was obtained from the application of the entropy balancing method itself.
Do extended school day programs affect performance in college admission tests?

Table A-3. Estimated Effects (with Weighting) of the Extended School Day Program on College Admission Test (ENEM) Scores – DD Specification Combined with Entropy Balancing Method.

|                     | Full-Time Program |                     |                     |                     |                     |
|---------------------|-------------------|---------------------|---------------------|---------------------|---------------------|
|                     | Math (1)          | Sciences (2)        | Humanities (3)      | Language (4)        | Writing (5)         |
| Treatment           | 0.203***          | 0.214***            | 0.178***            | 0.201***            | 0.335***            |
|                     | (0.0259)          | (0.0320)            | (0.0295)            | (0.0315)            | (0.0291)            |
| Observations        | 254,025           | 257,912             | 261,422             | 257,535             | 290,742             |
| R-squared           | 0.074             | 0.073               | 0.076               | 0.086               | 0.084               |
| School/Year FE      | Yes               | Yes                 | Yes                 | Yes                 | Yes                 |

|                     | Quasi-Full-Time Program |                     |                     |                     |                     |
|---------------------|-------------------------|---------------------|---------------------|---------------------|---------------------|
|                     | Math (6)                | Sciences (7)        | Humanities (8)      | Language (9)        | Writing (10)        |
| Treatment           | 0.0924***               | 0.0999***           | 0.0985***           | 0.108***            | 0.166***            |
|                     | (0.0196)                | (0.0218)            | (0.0198)            | (0.0193)            | (0.0230)            |
| Observations        | 269,437                | 273,589             | 277,274             | 273,122             | 310,794             |
| R-squared           | 0.064                  | 0.063               | 0.069               | 0.079               | 0.069               |
| School/Year FE      | Yes                    | Yes                 | Yes                 | Yes                 | Yes                 |

Notes: The table presents estimates obtained from the combination of the DD specification of equation (6) and the application of the entropy balancing method. Observations are weighted according to the entropy balancing method. All columns include fixed effects for school and time. Standard errors are clustered at the school level.

***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.
Figure A-8. Event-Study Estimates (with Weighting) of the Effects of the Extended School Day Program on College Admission Test (ENEM) Scores – Event-Study Specification Combined with Entropy Balancing Method.

Notes. This figure shows plots of event-study estimates with weighting and corresponding 95-percent confidence intervals. Observations are weighted according to the entropy balancing method. Standard errors are clustered at the school level.