Long-term effects of a public school management program

Ferman, Bruno and Finamor, Lucas and Portela Souza, André and Silva Filho, Geraldo

April 2025

Long-term effects of a public school management program*

Bruno Ferman§, Lucas Finamor§, André Portela Souza§, and Geraldo Silva Filho¶

§Sao Paulo School of Economics, FGV
¶Brazilian Ministry of Social Development and Fight against Hunger

Abstract

While a substantial body of research documents a positive correlation between school management practices and student outcomes, experimental evidence on their causal effects — particularly over the long term — remains limited. We exploit the randomized implementation of "Jovem de Futuro," a management program implemented in Brazilian public high schools. Drawing on rich administrative data, we follow students' educational and labor market trajectories over fifteen years. We find short-term improvements in test scores and high school completion. In the long run, however, the program yields null to modest effects on college enrollment and graduation, labor market participation, and earnings.

I. INTRODUCTION

Improving educational outcomes remains a critical challenge in public education, particularly in settings where resource constraints and systemic inefficiencies prevail. One aspect that has received increasing attention is the role of leadership and management skills of school principals and pedagogical coordinators. There is substantial evidence showing that effective management practices are positively correlated with improved student outcomes (e.g., Grissom and Loeb (2011), Branch et al. (2012), Coelli and Green (2012), Bloom et al. (2015), Di Liberto et al. (2015), Grissom et al. (2021)). These papers find a great dispersion

*We would like to thank the Instituto Unibanco staff for sharing with us the data from the experiment and INEP staff for cooperation with their data. We are also in debt with Ricardo Paes de Barros, Ricardo Madeira, Vladimir Ponczek, Cecilia Machado, Amanda Arabage, and seminar participants at EESP-FGV, VIII CAEN-EPGE, IFS, LSE-CEP, and Instituto Unibanco for their comments and suggestions. Valério Londe provided excellent research assistance. An initial draft of this paper was part of Lucas Finamor's M.A. thesis at the Sao Paulo School of Economics - FGV, under the title "Quality of education and human capital decisions: experimental evidence from Brazil".

of managerial skills and practices of school principals and, usually, large positive correlations with student outcomes. However, experimental evidence on the causal effect of management practices on student outcomes remains limited, with mixed results, primarily focusing on short-term outcomes such as test scores (e.g., Fryer et al. (2017), Henriques et al. (2020), Rosa (2015)). Whether schools' managerial practices affect students' long-term outcomes remains an open question.

In this paper, we fill this gap by providing one of the first experimental assessments that examines the long-term effects of a management program in high schools. Our study exploits the randomized rollout of the "Jovem de Futuro" program in Brazilian public schools, which aims to enhance the leadership and management skills of school principals and pedagogical coordinators. We combine four administrative microdata sources, which yield a 15-year panel of individuals' trajectories, including information from secondary and tertiary education, national standardized exams, and labor market outcomes. This impressive dataset, combined with the experimental implementation of the program, allows us to investigate the effects on students' high school progression and graduation, test scores on national exams, college enrollment and completion, formal labor market participation and earnings, for as long as twelve years after high school graduation.

We find that the "Jovem de Futuro" program had a positive effect on students' test scores — consistent with previous analyses by Henriques et al. (2020) and Rosa (2015). Students from treated schools had average effects of 7.5% of a standard deviation across all subjects of the national exam. The program also increased on-time graduation and ever graduating from high school by 4–6% (marginally significant). However, we do not find significant effects on overall college enrollment and college completion, nor on subsequent formal labor market outcomes. The exception is a positive effect on enrollment in public flagship universities and in selective majors of between 12% and 16%.

We use the relationship between the control group test scores and their outcomes on college enrollment to rationalize these findings. The magnitude of the test score improvements appears insufficient to generate meaningful effects on college enrollment and subsequent outcomes. Larger effects on overall college enrollment would require larger effects on test scores. This analysis also sheds some light on why we find results for selective majors. As the relationship between enrolling in selective majors and test scores is more convex, small gains in proficiency can push some individuals towards these institutions and majors. Therefore, the link between short-term test score gains and long-term outcomes varies significantly depending on the specific outcome and can often be highly nonlinear.

This article contributes to several literatures. We contribute to the literature on school management (Grissom and Loeb, 2011, Branch et al., 2012, Coelli and Green, 2012, Bloom et al., 2015, Di Liberto et al., 2015, Rosa, 2015, Fryer et al., 2017, Henriques et al., 2020, Grissom et al., 2021) by providing the first experimental evidence on the long-term effects of a management program in public schools. We also contribute to the literature analyzing the long-term effects of education and different schools' characteristics. In particular, on whether short-term effects on test scores are good predictors of long-term gains (Jackson, 2010, Deming, 2011, Deming et al., 2014, Angrist et al., 2016, Sass et al., 2016, Lavy, 2020, Beuermann et al., 2023, Jackson et al., 2024). We show how short-term test score gains do not translate to overall college enrollment and attainment, but they may generate effects on selective majors. This is an important finding, as it shows that the relationship between short-term and long-term outcomes is outcome-dependent and may not be linear.

The remainder of the paper is organized as follows. Section II provides an overview of the "Jovem de Futuro" program and its institutional context. Section III describes the data sources, while Section IV presents the empirical strategy employed in the study. Section V presents our main results and robustness checks. Section VI discusses the main findings, rationalizing our observation of positive effects on test scores alongside mostly null effects on subsequent outcomes. Finally, Section VII concludes.

II. INSTITUTIONAL SETTING

II.1 EDUCATIONAL SYSTEM IN BRAZIL

Basic education in Brazil is divided into two large tiers: primary education (grades 1 to 9, students between 6 and 14 years old) and secondary education (grades 10 to 12, students between 15 and 17 years old). Therefore, the expected duration of high school is three years. Most students (around 88%) are enrolled in public schools.[1] Public schools are tuition-free and, on average, of lower quality. For instance, taking the national standardized exam at the end of high school in 2008, the private-public overall gap in performance was 2.08 standard deviations.

Post-secondary education is also offered by a mix of public and private institutions. In contrast with basic education, public universities are large, prestigious, tuition-free, and, therefore, very selective institutions. Private institutions charge tuition, and, in only a

[1]Brazilian Education Census 2008, 87.95% of students in basic education were enrolled in public schools.

minority of them, few students can be awarded scholarships or student loans. In general, private colleges have a lower reputation and are less attractive than their public counterparts. In fact, Binelli et al. (2008) computed that, in 2003, there were, on average, nine applicants for each place in public universities, while there were only 1.5 applicants for each place in private colleges.

Admission is college-major specific. Students must choose a major before attending classes. Therefore, selectiveness varies not only among colleges but also across majors. Before 2010, each university had its own admission system, which could include a university-specific entrance exam. In 2010, a centralized system was introduced to unify entrance exams using the standardized national exam. Adoption of the centralized system is voluntary.[2]

Affirmative actions in Brazil started in 2002 in a few public universities, mainly using quotas that benefited black students and students from public high schools. Adoption of affirmative action increased over time, and in 2012, a federal law was approved, the "Quotas Law", which establishes that 50% of all places in public universities administered by the federal government should be reserved for students from public high schools.[3]

II.2 THE "JOVEM DE FUTURO" PROGRAM

Jovem de Futuro is an initiative implemented by Instituto Unibanco focusing on improving management at public high schools in Brazil.[4] The program aims at *"(...) strengthening the leadership and management skills of school principals and pedagogical coordinators, supporting them with data, indicators, goals, processes, training, advice, and various materials. The objective has always been to increase the retention of all students in school and the high school completion rate, with higher levels of learning."*(translated from Henriques et al. (2020), page 11). The program was implemented directly in the participating schools. They were offered training and assistance to develop, execute, and monitor a strategic plan. School staff were introduced to several management technologies (such as the logic model and guidance to adopt output-oriented management) and tools to foster the participation of students, their parents, and school staff. Agents from the Instituto Unibanco supervised the implementation of these new technologies and monitored schools on a weekly basis. Lastly, schools received additional

[2]For more details, check Machado and Szerman (2021) and Barahona et al. (2023).
[3]For more details, check Barahona et al. (2023).
[4]The Instituto Unibanco (*Unibanco Institute*) is a private organization founded in 1982 that aims to improve the quality of public education in Brazil.

resources equal to BRL 100 (USD 48) per student per year.[5] The goal of the transfer of resources was to help schools execute their established plans. Schools were continuously monitored during the three-year evaluation window, and the program could be discontinued if schools did not adhere to their own plans and goals (Henriques et al., 2020).

The program has been implemented since 2008, and a significant fraction of schools were part of an experimental evaluation of the program. The experimental window is three years long, which is also the expected duration of high school. That is, the students enrolled in the first year of high school when the program started are expected to receive the complete treatment for the three years they are enrolled in high school. There are three distinct waves of the program. The first wave was implemented in 2008–2010 and 2010–2012. The second wave was in 2013–2015 and 2014–2016; the last wave started its implementation in 2015. We focus here on the first wave to have a longer time window to analyze results, and because the second and third waves had shorter experimental windows, which precludes us from having a fully-treated and a pure control cohort for the entire (expected) duration of high school.

The first part of the program was implemented in the states of Minas Gerais (MG) and Rio Grande do Sul (RS). The treated schools received the program in 2008–2010, while the control group received the treatment immediately after the treatment group in 2011–2013. The subsequent phase selected schools from Rio de Janeiro (RJ) and São Paulo (SP), starting the program in 2010 for the treated schools and in 2013 for the control group. Importantly, since most high school students who began in 2008 would no longer be in high school by 2011, we do not expect significant contamination from the fact that control schools started receiving treatment in 2011.[6] At each locality, public schools could freely subscribe to the program, knowing in advance the program's structure and the experimental design. The assignment to treatment and control was completely random, using pairwise randomization, with stratification using school localization, number of students, proportion of students in high school, and baseline proficiency.[7] The number of schools in each group and area is presented in Table 1 below. There are two areas in the state of São Paulo, the metropolitan region and the Vale do Paraiba region, each with 40 schools.

[5]The rule that determined the amount of resources had a slight change for schools in 2008–2010, compared to those in 2010–2012. In the first group, the number of students considered was the total number of students in school; meanwhile, in the second phase, only the number of students in high school was considered, but a minimum transfer of R$ 100,000 yearly was guaranteed for all schools. For details, please check Rosa (2015).

[6]In fact, among the students enrolled in the first year of high school in control schools, only 12.2%, 3.8%, and 1.5% were enrolled in the same school in years 4, 5, and 6, when these control schools implemented the program. Figure A.1 shows the proportion of students enrolled in the original school by year and type of school.

[7]In Minas Gerais, the randomization procedure had minor differences. The number of control schools was

Table 1: Treatment Assignment

Area	Start Year	# Strata	# Schools		
			Total	Treated	Control
MG	2008	4	48	20	28
RS	2008	25	50	25	25
RJ	2010	15	30	15	15
SP1	2010	20	40	20	20
SP2	2010	20	40	20	20
Total	-	84	208	100	108

Notes: The table shows for each of the five areas the start year of the evaluation window, the number of strata, the total number of schools, and the number of treated and control schools. The areas comprise schools from the state of Minas Gerais, Rio Grande do Sul, Rio de Janeiro, and two areas from the state of São Paulo (metropolitan region and Vale do Paraíba.)

Table 2 presents the balance between treatment and control groups for several baseline measures, including school infrastructure, characteristics, and performance. As expected, for all variables except for the internet connection indicator, we cannot reject the null hypothesis that treated and control schools have the same mean. One point of attention is the baseline score using the national standardized exam (ENEM), which shows that the treated schools have higher average scores. Despite this difference not being statistically significant, we will control for baseline scores in our main specification.

The randomization guarantees the internal validity of the experiment. The exercise assessing the baseline balance on important variables also provides supporting evidence of the comparability of treatment and control schools. However, to help interpret the results, it is also important to understand which schools are part of the experiment. Figure 1 shows the standardized scores of high schools on the national exam (ENEM) the year before the program started. The gray dots are all schools in the participating states, and the red circles show the position of the schools participating in the experiment in the distribution of scores. We can see that schools are dispersed along the entire distribution. This, added to the fact that the experiment takes place in four different states, 36 municipalities, and 208 schools, gives substantial external validity to our results.

larger than that of treatment schools; hence, strata with few schools were created instead of pairs.

Table 2: Baseline balance

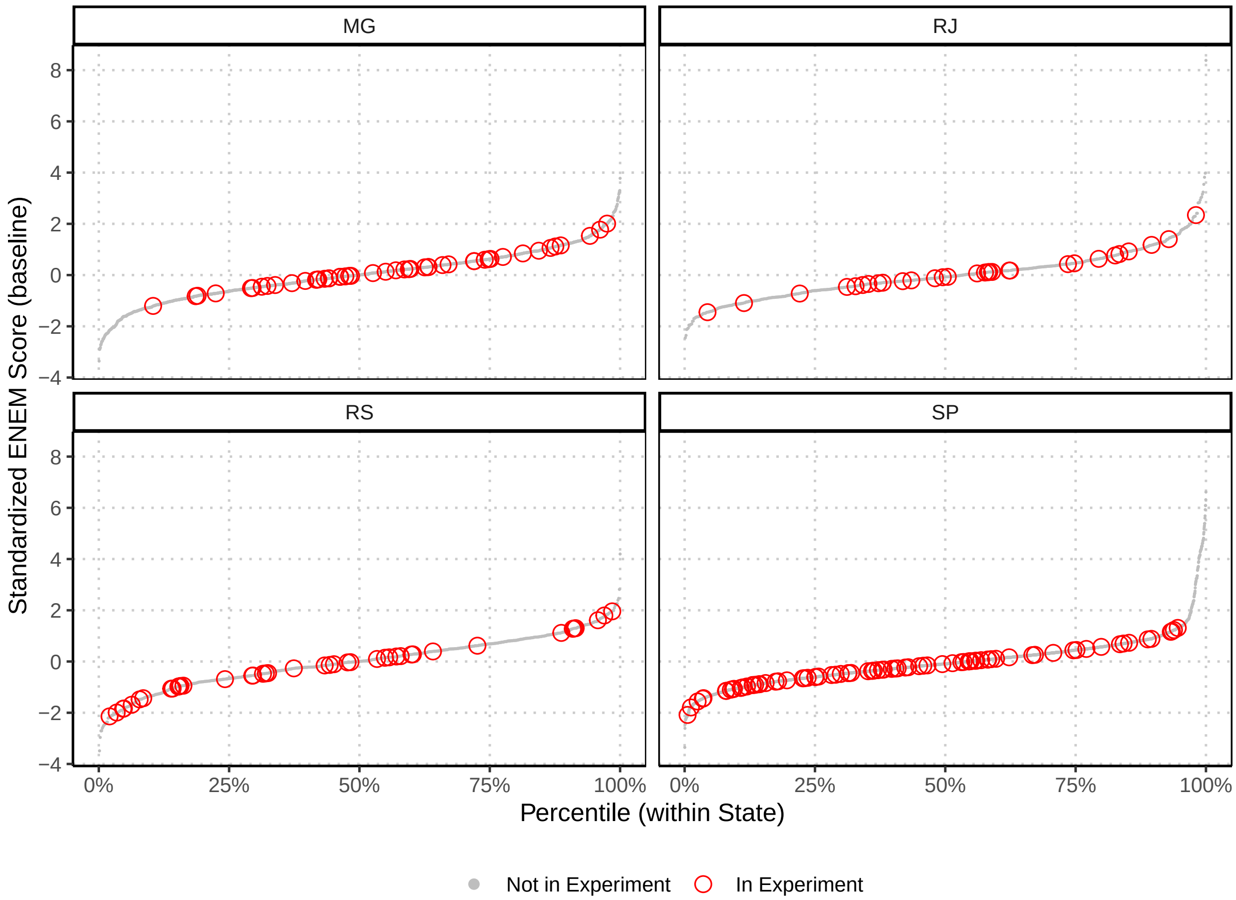
Variable	Mean Treated	Mean Control	Diff	P-value
# PCs	14.662	15.419	-0.638	0.536
# Classrooms	18.280	18.694	-0.477	0.675
Has Internet	0.880	0.991	-0.118	0.005
# Staff	105.360	96.787	10.748	0.394
# Students	1577.450	1625.815	-40.944	0.519
# Students - High School	864.200	914.556	-32.547	0.454
# Students - 1st year High School	377.510	392.287	-9.071	0.665
# Groups	44.730	44.648	0.195	0.907
# Groups - High School	23.630	24.241	-0.189	0.874
# Groups - 1st year High School	10.100	9.981	0.268	0.637
Prop taking ENEM	0.474	0.463	0.019	0.312
ENEM score (std)	-0.123	-0.189	0.039	0.427

Notes: The table shows for each variable (each row), the average for the treated schools (first column), the average value for control schools (second column), the regression-adjusted difference (third column), and the p-value for the statistical test whether the estimated difference is zero. The regression adjustment includes strata fixed effects. Standard errors are clustered at the strata level. All variables are measured in the year before the intervention started.

III. DATA

This paper uses a collection of administrative data, including school records for secondary and tertiary education, national exams, and matched employer-employee data. Our starting point is the Brazilian Education Census, which has information on the universe of students enrolled in primary and secondary education in any school in the country. For every student-year observation, it is possible to establish the student's enrollment status, grade, and school. The microdata is available between 2007 and 2022. Using this dataset, we can generate the list of first-year high school students in the initial year of the intervention at each participant school. All students in this list are considered experiment participants with treatment status defined by the school status in the randomization. We keep enrolled students who have completed the first year of study, whether approved or not. If students change schools or even drop out after the first year, we keep them in the sample with their initial treatment status.

Figure 1: Distribution of schools by baseline ENEM score



Notes: The graph shows the position of each school and their associated score in the standardized ENEM score in the baseline year. The scores were standardized considering all public high schools within the same state (excluding the schools in the experiment). Each plot shows the schools in a given state (MG for Minas Gerais, RJ for Rio de Janeiro, RS for Rio Grande do Sul, and SP for São Paulo). The gray dots show all public high schools that were not in the experiment, while the red circles show the schools in the experiment.

We also use the Education Census to track the students' progress through high school, using all the available data until 2022.

At the end of high school, students can take a National Exam called "National Exam of Upper Secondary Education" (ENEM). ENEM has had four objective exams since 2009 based on the Item Response Theory (IRT): natural sciences, human sciences, language, and mathematics, as well as a writing exam. The exam was created in 1998 to evaluate the quality of high schools in Brazil. Over time, the exam started to be used as a part of the college admission process and culminated as the main process nowadays to select students for public universities and beneficiaries for public scholarships, loans, and academic exchanges.

Students beyond the typical age for high school completion can also obtain a high school diploma if they achieve a minimum score on the exam. Microdata for this exam is available between 2009 and 2022.

The Higher Education Census is similar to the Education Census in that it covers the universe of students enrolled in tertiary programs. The dataset allows us to observe students' enrollment status, including major and college. Microdata is available between 2009 and 2022.

Lastly, we use the "Annual List of Social Information" (RAIS) dataset, which contains administrative records of all formal labor contracts in the country. Every observation is a formal labor contract between a worker and a firm in a given year, which specifies the contract status at the end of the year (active or inactive), the duration of the work agreement, the wage, and the number of contracted hours. RAIS is available between 2007 and 2022.

As the four datasets have universal coverage, we can follow students from their first year in high school until 2022 (10–12 years after expected high school graduation). We create a panel at the student level. For every year, we have information on whether this student was enrolled in high school or tertiary education, took the national exam (and their scores), and their employment status. One caveat is that the Education census uses a student ID as an identifier, while the other three datasets use the individual tax identifier number (CPF). Using the available administrative data, we identify CPFs for 80% of our sample. In Section V.6, we show how this attrition is the same for treated and control schools and how effects do not vary across the sample with and without CPFs for the outcomes that do not depend on having this information.

IV. EMPIRICAL STRATEGY

Our empirical strategy relies on exploiting the program's random assignment. We estimate the causal effects of the program using the OLS estimator of the following equation:

$$Y_i = \alpha + \beta treat_{s(i)} + \eta_{g(i)} + \gamma X_i + \delta Z_{s(i)} + \varepsilon_i \tag{1}$$

Where Y_i is the outcome of interest for student i. Each student belongs to a school $s(i)$, which participated in the randomization procedure in strata-group $g(i)$. Schools have different sizes, and the proportion of treated and control schools differs in some regions, so we include strata fixed effects ($\eta_{g(i)}$). β is our coefficient of interest, capturing the causal effect of the program. We keep all schools with their original assignment, and the same goes for students. That is, even if they change schools after the program started, we keep their original assignments to treatment or control, depending exclusively on their original assignments. Therefore, our strategy estimates the intention-to-treat (ITT) effect.

To increase precision, we add demographic controls at the student level (gender, age, and race) and school controls (baseline score at ENEM and average outcome for a previous cohort). As expected, the demographic controls barely change the point estimates and reduce the estimated standard errors. Introducing baseline ENEM scores reduces point estimates for most outcomes. That is likely because treated schools had higher initial proficiency (even though the difference is not statistically significant). We also add the school average outcome for a previous cohort to increase precision, as it is predictive of the outcome of interest. To be conservative, we show results including this variable. Results without these controls are also available. Following the recommendations of De Chaisemartin and Ramirez-Cuellar (2024), we clustered our standard errors at the strata level. We use the procedure suggested by Ferman (2019) to assess how inference is implemented, and we do not find any evidence of over- or under-rejection.

V. RESULTS

V.1 HIGH SCHOOL PROGRESSION

We start by analyzing how students progressed in high school in response to the program. Table 3 shows the estimates for the treatment effect of the program on students progressing through the high school years. In the first column, we can see that the estimated treatment effect for students being approved in the first year of high school is 0.009 percentage points (pp), non-statistically significant (p-value 0.64). The interpretation of this number is that in the treated schools, 0.9 more students out of 100 were approved in the first year in the expected year, that is, the first year of intervention. The estimated points rise to 1.9pp and 2.4pp in the subsequent two years. That is unsurprising, as the program was cumulatively implemented in these three initial years.

In the third column, we show the effect of the program on students graduating from high school in the expected year (three years after starting high school). On-time graduation was 2.4pp higher in the treatment group (p-value 0.087). Taking into account the control mean of on-time graduation (41%), this represents an increase of approximately 6%. In the fourth column we use as outcome whether the student graduated from high school at any point

Table 3: Treatment Effects - High School progression

	Approved 1st HS Expected Year	Approved 2nd HS Expected Year	Graduate HS Expected Year	Graduate HS Any Year	Graduate HS or EJA Any Year
	(1)	(2)	(3)	(4)	(5)
Treat	0.009	0.019	0.024	0.020	0.016
(s.e.)	(0.020)	(0.018)	(0.014)	(0.011)	(0.009)
[p-value]	[0.645]	[0.286]	[0.087]	[0.074]	[0.066]
N Obs	65,435	65,435	65,435	65,435	65,435
N Schools	207	207	207	207	207
N Strata	84	84	84	84	84
Control Mean	0.678	0.490	0.410	0.503	0.604

Notes: The table shows the estimates of equation 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome. The first two is whether students were approved in the first and second year of high school, in the expected year. The third and fourth columns whether students graduated from HS in the expected year (third column) or ever (fourth column). The last column uses as outcome whether the individual graduated from high school or from the special education for adults at any point in time. The mean for the control group is also displayed.

(up to 10–12 years after expected graduation) and in the fifth column whether they also graduated from a special education program dedicated to the adult population. Therefore, we can see that the effect on on-time graduation comes partially from inducing individuals who would graduate late or from the special education program to graduate from regular school on time, and partially from inducing individuals who would not have graduated otherwise to do so.

V.2 STANDARDIZED NATIONAL EXAM (ENEM)

At the end of high school, students can take the standardized national exam, which can be used, among other uses, for college admission. Table 4 shows estimates for the effect of the PJF program on exam take-up and performance. The first two columns show the estimates for take-up in the expected year (column 1) or any year. Students from treated schools are 1.9pp and 2.3pp more likely to take the exam. This represents an increase of 4-6%, considering the control mean.

Table 4: Treatment Effects - ENEM

Outcome:	Takeup	Takeup	Math	Language	Science	Humanities	Essay
Year:	Expected	Any	Expected	Expected	Expected	Expected	Expected
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.019	0.023	0.092	0.100	0.049	0.059	-0.001
(s.e.)	(0.011)	(0.011)	(0.039)	(0.032)	(0.030)	(0.026)	(0.020)
[p-value]	[0.092]	[0.039]	[0.020]	[0.002]	[0.100]	[0.027]	[0.964]
N Obs	52,545	52,545	15,986	15,986	16,365	16,365	15,991
N Schools	207	207	207	207	207	207	207
N Strata	84	84	84	84	84	84	84
Control Mean	0.299	0.499	-	-	-	-	-

Notes: The table shows the estimates of equation 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome. The first two columns use as outcome variables whether the student sit on the exam in the expected year (first column) or in any year (second column). The next five columns show effects on performance in the exam, using standardized scores for math, language, science, humanities, and essay. The performance variables use only data from students taking the exam at the expected year.

The following five columns show the program's effect on performance in the four objective parts of the exam (math, language, natural sciences, and humanities) and essay. We see significant positive effects for all four objective exams, but not for writing. The outcomes are standardized to have a mean of zero and a standard deviation of one in the control group. The estimates are sizable, ranging from 4.9-10% of a standard deviation. These scores are considered solely for students who took the exam in the expected year (year of expected graduation from high school). As the first column shows, the program increases exam take-up, so these estimates cannot be read directly as the causal effect of the program on test scores. However, we believe they can be seen as lower bounds for the effects, as it is more likely that the program induced students with relatively lower proficiency to sit on the exam.

These results are also in line with previous evaluations of the program. Silva (2010), Barros et al. (2012), and Oliva (2014) present the effects using an exam created and annually implemented by the Instituto Unibanco in the participant schools as a measure of students ' proficiency. The estimated effects are positive in all studies for all years and areas analyzed. Rosa (2015) uses ENEM to assess the program's effects on students' proficiency. As the author uses the non-identified version of the data, he can only assess the effects for students who take the exam and are enrolled in the last year of high school in the participating schools.

The results are similar to ours, with an average effect on language and math scores around 15% of a standard deviation. Our estimates are more conservative and can better approximate the ITT effect as we use the sample of students enrolled in the initial year of the program, regardless of their future educational trajectory.

V.3 COLLEGE ACCESS

We now turn to enrollment in college. Table 5 shows the results for students enrolling in any period until 2022, 12 years after expected high school graduation for those in RS and MG, and 10 years for those in SP and RJ. In the first column, we see that about 39.9% of the students in the control group enroll in a tertiary course. The point estimate for the effect of the program is positive and non-statistically significant, at 1.1pp. In the next two columns, we break this by enrollment at public or private universities[8]. For public universities, we have an increase of 0.5pp, which is marginally significant (p-value 0.100). Even though this coefficient seems small, it is a large effect compared with the control group, as only 4% of the students in control schools ever enroll in a public university. This is important as, as discussed in Section II.1, public universities are the most prestigious and selective institutions in Brazil.

In the last column, we show the effects of students ever enrolling in a course through affirmative action policies. As discussed in Section II.1, around this period, many public institutions started adopting affirmative action policies, particularly using quotas for students in public high schools and from minority backgrounds. The program significantly increased the share of students admitted through these policies—an increase of 0.4pp on a basis of 2.3%. As most of the affirmative action policies were adopted by public institutions, and the point estimates are almost the same, it is likely that the increase in enrollment in public universities came from students in treated schools through the affirmative action policy.

The evidence above suggests that students are more likely to be admitted using affirmative action and to be enrolled in public universities. It is then informative to evaluate how good this placement is in terms of college-major quality. We follow Moreira (2019) and construct a measure of quality and selectivity of majors using entrants' ENEM test scores. For each course, we use the average score of first-year students in all four objective parts of ENEM. Since a student can take the exam in several years, we only used the last score available prior

[8]The same student could be enrolled in public and private universities in this 12-year window, which is why public and private do not sum to the total enrollment.

Table 5: Treatment Effects - College Enrollment

	Ever Enrolled			
	in College	in Public Universities	in Private Universities	with Affirmative Action
	(1)	(2)	(3)	(4)
Treat	0.011	0.005	0.008	0.004
(s.e.)	(0.011)	(0.003)	(0.010)	(0.002)
[p-value]	[0.307]	[0.100]	[0.437]	[0.046]
N Obs	52,545	52,545	52,545	52,545
N Schools	207	207	207	207
N Strata	84	84	84	84
Control Mean	0.399	0.040	0.377	0.023

Notes: The table shows the estimates of equation 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome. Outcomes are indicators for students ever enrolling in any tertiary program (column 1), in a public university (column 2), in a private institute (column 3), and being admitted using affirmative action policies (column 4).

to the university admission[9]. Therefore, for every college-major-year we have an average score, and we rank all majors in Brazil based on this score. We will use the percentile of this distribution as a measure of major quality and selectiveness. In order to use representative measures, we eliminate all majors in which only less than 10% of students have available grades.[10]

Table 6 presents the treatment effect on the probability of a student being admitted to a major in several percentiles of the quality-selectiveness distribution. Since we already know that treatment impacted positively ENEM's scores and the measure is comprised exclusively of these grades, we use a lag of the quality measure in order to avoid a mechanical effect on increasing the majors' selectiveness. Splitting the courses below or above the median, we see similar point estimates (0.8pp and 0.6pp), both non-statistically significant. Analyzing more selective majors, those in the top 30%, 20%, and 10%, we still see positive estimates, which are large in proportional terms. For instance, there is an increase of 0.9pp in the probability

[9]We only use ENEM data starting in 2009.

[10]We choose not to use ENADE (Student Proficiency National Exam) as a measure of quality because of (i) the unavailability of annual data, (ii) the lack of evaluation for several courses, and (iii) the lack of incentives for students to sit the exam.

Table 6: Treatment Effects - College Enrollment (Quality)

	Ever Enrolled by percentile				
	Below median	Above median	Above 70th	Above 80th	Above 90th
	(1)	(2)	(3)	(4)	
Treat	0.008	0.006	0.006	0.009	0.004
(s.e.)	(0.007)	(0.009)	(0.007)	(0.005)	(0.002)
[p-value]	[0.294]	[0.516]	[0.378]	[0.075]	[0.128]
N Obs	52,545	52,545	52,545	52,545	52,545
N Schools	207	207	207	207	207
N Strata	84	84	84	84	84
Control Mean	0.298	0.210	0.113	0.062	0.025

Notes: The table shows the estimates of equation 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome. Outcome variables are indicators for students enrolling in majors-universities at different parts of the distribution of program quality/selectiveness: below median (column 1), above median (column 2), in the top 30% (column 3), 20% (column 4), and 10% (column 5).

of students enrolling in the top 20% best courses, corresponding to an increase of 15%. This evidence eases the concern that students enrolled in more prestigious universities but in less demanded majors.

Lastly, Table 7 assesses whether treated students were more likely to graduate from a tertiary course 6, 8, and 10 years after expected high school completion. We see minimal positive point estimates, which are non-statistically significant. It is worth noting how low the graduation rates are for the control group. Ten years after (expected) higher school completion, only 14.9% of students graduated from college.

All together, these results suggest two possible mechanisms. One is that there is no effect on college enrollment, although there is an effect on college quality composition. Alternatively, there is a positive impact on college enrollment, together with an increase in college dropouts among treated students.

Table 7: Treatment Effects - College Graduation (cumulative)

	Graduated from college		
Years after expected HS graduation	6	8	10
	(1)	(2)	(3)
Treat	0.002	0.001	0.003
(s.e.)	(0.003)	(0.005)	(0.007)
[p-value]	[0.533]	[0.810]	[0.630]
N Obs	52,545	52,545	52,545
N Schools	207	207	207
N Strata	84	84	84
Control Mean	0.062	0.110	0.149

Notes: The table shows the estimates of equation 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome. Outcome variables are indicators for whether the students have already graduated from a tertiary program six (column 1), eight (column 2), and ten (column 4) years after expected high school graduation.

V.4 LABOR MARKET

Using data from RAIS, we can investigate labor market outcomes for these individuals. Table 8 shows the results for formal employment (indicator) and for log-wages for those employed in the final years of your sample, between 7 and 10 years after expected high school completion. We do not see any effects on formal employment. The coefficients are extremely small and switch signs. For the final year, we can rule out increases of more than 1pp, on the basis of 47.3%. For log-wages, our point estimates are negative and mostly non-significant, except for the final year. Therefore, we do not see evidence of improved formal labor market outcomes in the 10 years following high school.[11]

[11] The period we analyze is characterized by an increase in the unemployment rate and a reduction in formalization, which had a larger effect on younger workers (Corseuil et al., 2021). Therefore, some of our results in the labor market may be affected by these particular economic conditions. This reinforces the advantage of a randomized protocol, which guarantees internal validity of our empirical strategy, since treated and control students are comparable under any economic environment.

Table 8: Treatment Effects - Labor Market

Years after expected HS graduation	7	8	9	10
	(1)	(2)	(3)	(4)
Panel A. Outcome: Formal Employment				
Treat	−0.003	−0.006	0.000	0.001
(s.e.)	(0.006)	(0.006)	(0.005)	(0.005)
[p-value]	[0.651]	[0.321]	[0.985]	[0.858]
N Obs	52,545	52,545	52,545	52,545
N Schools	207	207	207	207
N Strata	84	84	84	84
Control Mean	0.465	0.458	0.460	0.473
Panel B. Outcome: Log-Wage				
Treat	−0.004	−0.001	−0.005	−0.017
(s.e.)	(0.005)	(0.006)	(0.006)	(0.008)
[p-value]	[0.413]	[0.916]	[0.419]	[0.043]
N Obs	22,993	22,585	22,805	23,297
N Schools	207	207	207	207
N Strata	84	84	84	84
Control Mean	7.396	7.352	7.370	7.430

Notes: The table shows the estimates of equation 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome. In panel A, we use as outcome variable an indicator of having a formal labor market contract in a given year (7–10 years after expected high school graduation). In panel B, we use as outcome the monthly log-wages for those employed. We trim the bottom and top 1% of log-wages.

V.5 HETEROGENEITY

We investigate whether the effects presented so far may mask some heterogeneity at the individual level (by gender and race) or the school level (by terciles of baseline proficiency). For the baseline school proficiency, we use the school average ENEM score in the year before the intervention. We classify each stratum as belonging to the bottom, middle, or top tercile of the ENEM distribution in each state. Note that as baseline proficiency scores were used as one of the variables for the stratification, schools in the same strata are very similar in terms of baseline performance. As the sample size is not very large, we see these results as mostly indicative, as we are not powered enough to conduct a thorough analysis. Table 9 presents the results for some selected outcomes: high school graduation in any year, ENEM take-up (any year), ENEM math score, tertiary enrollment (any, in public universities, and majors in the top 10%), college graduation (10 years after expected high school completion), and formal employment (10 years after expected high school completion).

Table 9: Treatment Effects - Heterogeneity

	HighSchool	ENEM		Enrollment			Coll Grad	Employment
	Graduation	Take-up	Math	Any	Public	≥90th	Year 10	Year 10
Panel A - By gender								
Treat x Woman	0.022	0.021	0.104	0.010	0.006	0.004	−0.002	0.010
	(0.011)	(0.012)	(0.046)	(0.013)	(0.003)	(0.003)	(0.008)	(0.008)
Treat x Men	0.017	0.028	0.084	0.017	0.004	0.004	0.012	−0.010
	(0.013)	(0.012)	(0.037)	(0.011)	(0.003)	(0.003)	(0.006)	(0.005)
P-value [Woman=Man]	0.583	0.498	0.542	0.531	0.714	0.840	0.037	0.055
Panel B - By race								
Treat x Black/Native	0.010	0.007	0.084	0.002	−0.001	−0.002	−0.008	0.007
	(0.016)	(0.015)	(0.066)	(0.017)	(0.004)	(0.003)	(0.009)	(0.011)
Treat x White/Asian	0.022	0.016	0.115	0.000	0.009	0.004	0.001	−0.001
	(0.015)	(0.015)	(0.042)	(0.015)	(0.005)	(0.004)	(0.011)	(0.009)
P-value [Black=White]	0.350	0.601	0.618	0.896	0.100	0.126	0.359	0.576
Panel C - By school quality (terciles)								
Treat x T1	0.019	0.027	0.075	0.002	0.003	0.002	0.002	−0.011
	(0.027)	(0.014)	(0.045)	(0.012)	(0.003)	(0.003)	(0.007)	(0.009)
Treat x T2	0.012	0.018	0.063	0.015	0.005	0.002	0.000	0.003
	(0.017)	(0.023)	(0.078)	(0.022)	(0.005)	(0.006)	(0.015)	(0.008)
Treat x T3	0.031	0.024	0.138	0.013	0.006	0.008	0.008	0.008
	(0.017)	(0.023)	(0.078)	(0.022)	(0.005)	(0.006)	(0.015)	(0.008)
P-value [T1=T2]	0.816	0.739	0.901	0.609	0.802	0.937	0.908	0.226
P-value [T1=T3]	0.704	0.852	0.297	0.553	0.607	0.204	0.573	0.107
P-value [T2=T3]	0.394	0.837	0.389	0.931	0.877	0.342	0.625	0.686

Notes: The table shows the estimates of equation 1 interacted with heterogeneity indicators, by gender (in Panel A), by race (in Panel B), by school quality tercile (in Panel C). We display the coefficient for the treatment indicator for each heterogeneity level, their corresponding standard errors (clustered at the strata level), and the p-value. Each column shows a different outcome. Respectively the outcomes are on-time high school graduation, on-time ENEM takeup and math scores, ever enrolling in any tertiary program, in a public university, or in a major-university in the top 10% of programs, college graduation measured ten years after expected high school graduation, and formal employment measured ten years after expected high school graduation.

In terms of gender, we do not see drastically different results apart from college graduation, which is higher for men, and employment, which is higher for women. Regarding race, we see larger estimates for White/Asian students than for Black/Natives—particularly high school graduation, ENEM take-up and scores, and tertiary enrollment in public and selective majors. As White/Asian students are, on average, from more affluent backgrounds, these results are consistent with students with better socioeconomic status exhibiting larger effects. We see similar results when exploring heterogeneity by school baseline proficiency. Schools from the top of the distribution consistently have larger estimates. These results provide suggestive evidence that the program may be more effective in schools that were already better in the baseline and for students from (relatively) more affluent backgrounds. However, we cannot reject the null hypothesis that these coefficients are the same, since when breaking the sample into smaller groups, we lose power.

V.6 ROBUSTNESS

Our empirical strategy directly exploits the random assignment of the treatment status. Therefore, we need very minimal assumptions to interpret most of the effects presented here as the causal (intention to treat) effect of the program. In Table 10, we show how the estimates change depending on the choices of included controls. As an example, the table shows the effects of the outcome on enrollment in public universities. In the first column, we add solely the strata and calendar year fixed effects. In the second column, we add control variables at the individual level (gender, age, and race). As expected, the point estimates barely change while we gain more precision. In the third column, we present our baseline specification, including the two controls at the school level (baseline ENEM score and the outcome variable for the cohort graduating before the program started). The point estimate halves, basically due to the inclusion of the baseline ENEM score. As we saw in the balance table, treated schools had, on average, baseline scores 4% higher (non-statistically significant). As ENEM scores are highly correlated with most outcomes, their inclusion reduces the treatment effect estimates for most outcomes. To be conservative, we adopt this as the baseline specification. Most results would be larger without these controls.

Apart from the high school progression outcomes, all remaining variables depend on the individual tax identifier number that is available for 80% of our sample. In Table A.1, we show how observing the individual identifier (CPF) is not correlated with the treatment (column 2) and also how the results for high school progression are similar when restricting the analysis to the sample for whom we have the identifier number (columns 3 and 5). We also removed individuals who dropped out of school in the first year of the program from the sample. Table A.1 also shows that there is no differential probability of dropping out by treatment status (column 1) and also how effects are qualitatively similar if we instead use

Table 10: Robustness - Different specifications

Outcome: College Enrollment: Public Universities			
	No Controls	Ind Controls	All Controls
Treat	0.010	0.009	0.005
(s.e.)	(0.004)	(0.004)	(0.003)
[p-value]	[0.019]	[0.017]	[0.100]
N Obs	52,545	52,545	52,545
N Schools	207	207	207
N Strata	84	84	84
Control Mean	0.040	0.040	0.040
Strata FE	✓	✓	✓
Ind. Controls	-	✓	✓
School Controls	-	-	✓

Notes: The table shows the estimates of 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column uses a different set of controls. In the first column, only strata and calendar year fixed effects are included. In the second column, we include individual-level controls (age, gender, and race). In the third column, we include school-level controls (baseline ENEM scores and the outcome variable for a previous cohort). The outcome variable for the three columns is ever enrolling in a public university.

the full sample (columns 3 and 4).

VI. DISCUSSION

The results discussed here show that the PJF program: (i) increased high school graduation, particularly on-time graduation, (ii) increased ENEM take-up and performance on the exam, (iii) had some effects on enrollment in public universities and selective majors, (iv) no effects on college graduation and, (v) no effects on the formal labor market. The results for general college enrollment and enrollment in private institutions were positive but not statistically significant. Lastly, the heterogeneity analysis showed suggestive evidence of larger effects for schools with higher baseline scores and students from more affluent backgrounds, using race as a proxy for socioeconomic status.

In order to better interpret these results, we analyze the relationship between the ENEM scores and the main college outcomes using the students in the control schools. Figure 2a shows average college enrollment by deciles of ENEM scores. For instance, students in the control group in the first decile of the ENEM score (conditional on taking the exam) have an average score of -1.4 standard deviation, and 25.5% are enrolled in college 4 years after expected high school graduation. For students in the top decile, their average score is 1.5 standard deviations, and 81.1% of them are enrolled. Panels 2b and 2c show the same graph for enrollment in public universities and the top 10% programs.

Together, these graphs help us interpret our results. Interestingly, total enrollment in year 4 shows a linear relationship with respect to ENEM scores. An increase of one standard deviation in ENEM scores is associated with a 20 percentage point increase in college enrollment. Using this relationship, we can predict the effect of the program from the increase in ENEM scores. The program increased the average ENEM score by 1/4(0.092+0.100+0.049+0.059)=0.075. As 52% of the individuals in the treated group take-up ENEM, the predicted increase in college enrollment is $(0.20 \times 0.075 \times 0.52) = 0.008$.[12] Our estimated coefficient for college enrollment in year 4 is 0.006. That is, in order to obtain larger effects on college enrollment, the program would need to have had larger effects on ENEM scores.

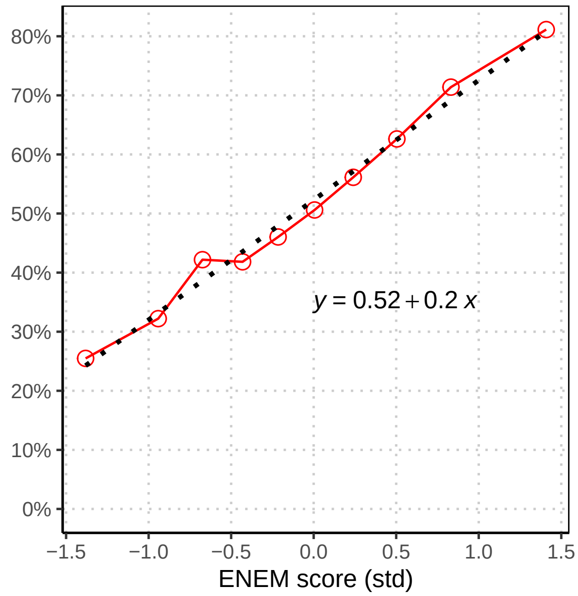
When we turn to the enrollment outcomes in public universities and selective majors, as expected, we see a steeper and convex relationship with ENEM scores. Enrollment in those universities and programs is concentrated at the top of the distribution. This is consistent with our findings that these effects were larger for schools with higher baseline proficiency and students with better socioeconomic status. These students and schools have higher ENEM scores, where smaller gains in proficiency could result in larger effects on enrollment in more selective programs. That is, the returns to ENEM scores could be very nonlinear, as the graphs show.

VII. CONCLUSION

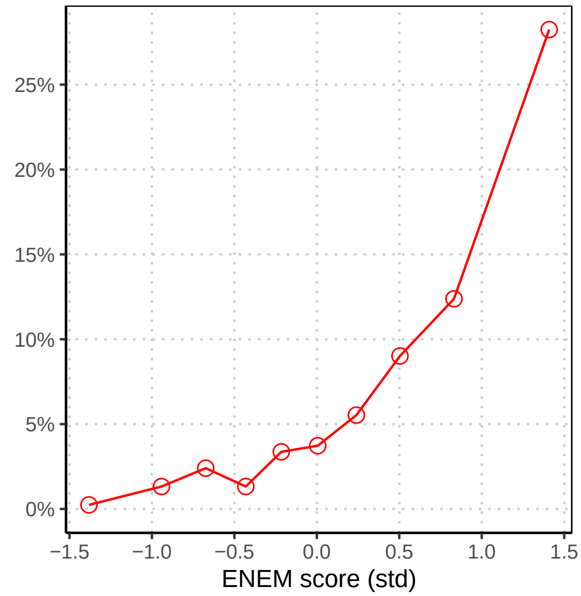
In this paper, we provide one of the first experimental evidence of the role of leadership and management skills of school principals and pedagogical coordinators on immediate student performance and longer term college and formal labor market outcomes. We leverage the randomized rollout of the Jovem do Futuro program in Brazilian public high schools that aimed to improve leadership and management skills of their managers, and the availability of high-quality administrative data to track students' educational and labor market trajectories.

[12] This result uses the average increase in ENEM score in the year of expected graduation, but the ENEM take-up in any year after high school completion.

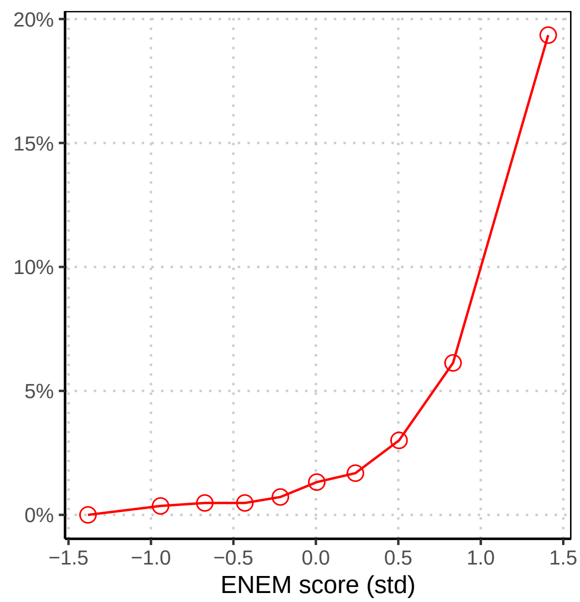
Figure 2: College enrollment variables by ENEM performance (control group)



(a) Total



(b) Public universities



(c) Majors in the top 10%

Notes: The graph shows the relationship between ENEM score and college outcomes for students in the control group. We show the average ENEM scores and average enrollment in college (panel a), enrollment in public universities (panel b), and enrollment in majors in the top 10% (panel c), by deciles of ENEM score. Each red circle plots these two averages for each decile. In panel (a) the dotted black line and numbers show the fitted OLS linear regression.

We find that the program had a positive effect on test scores of the range of 5-10% of a standard deviation. This is much smaller than the cross- and within-country correlations of the effects of management skills on student performance found in the literature. We find small effects on on-time high school graduation and enrollment in public universities and selective majors. However, we do not find significant effects on longer-term outcomes. There are no discernible effects on college admission, college graduation, and formal labor market employment.

Why there are no discernible effects in the longer term is an open question. We saw how increasing overall access to post-secondary education would likely require more substantial improvements in test scores. The null effects on college graduation could be driven by a combination of an increase in enrollment in public universities together with a differential delay in graduation caused by the deteriorated economic conditions, as presented by Finamor (2023). The muted effects in the labor market suggested that either these short-term gains vanished or these cognitive skill gains were not large enough to surpass the entry barriers to college or the formal labor market. Nonetheless, these gains could be significant for other domains in life. Unfortunately, despite being rich, the administrative data does not allow us to test this different hypothesis. These are possible routes for future research.

REFERENCES

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A. and Walters, C. R. (2016). Stand and deliver: Effects of boston's charter high schools on college preparation, entry, and choice, *Journal of Labor Economics* **34**(2): 275–318.

Barahona, N., Dobbin, C. and Otero, S. (2023). Affirmative action in centralized college admissions systems, *Unpublished, Working Paper* .

Barros, R., de Carvalho, M., Franco, S. and Rosalém, A. (2012). Impacto do projeto jovem de futuro., *Estudos em Avaliação Educacional* **23**(51): 214–226.

Beuermann, D. W., Jackson, C. K., Navarro-Sola, L. and Pardo, F. (2023). What is a good school, and can parents tell? evidence on the multidimensionality of school output, *The Review of Economic Studies* **90**(1): 65–101.

Binelli, C., Meghir, C. and Menezes-Filho, N. (2008). Education and wages in brazil. Working paper.

Bloom, N., Lemos, R., Sadun, R. and Van Reenen, J. (2015). Does management matter in schools?, *The Economic Journal* **125**(584): 647–674.

Branch, G. F., Hanushek, E. A. and Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals, *Technical report*, National Bureau of Economic Research.

Coelli, M. and Green, D. A. (2012). Leadership effects: School principals and student outcomes, *Economics of Education Review* **31**(1): 92–109.

Corseuil, C. H. L., Franca, M. P. and Poloponsky, K. (2021). A inserção dos jovens brasileiros no mercado de trabalho num contexto de recessão, *Novos estudos CEBRAP* **39**: 501–520.

De Chaisemartin, C. and Ramirez-Cuellar, J. (2024). At what level should one cluster standard errors in paired and small-strata experiments?, *American Economic Journal: Applied Economics* **16**(1): 193–212.

Deming, D. J. (2011). Better schools, less crime?, *The Quarterly Journal of Economics* **126**(4): 2063–2115.

Deming, D. J., Hastings, J. S., Kane, T. J. and Staiger, D. O. (2014). School choice, school quality, and postsecondary attainment, *The American economic review* **104**(3): 991–1013.

Di Liberto, A., Schivardi, F. and Sulis, G. (2015). Managerial practices and student performance, *Economic Policy* **30**(84): 683–728.

Ferman, B. (2019). Assessing inference methods, *arXiv preprint arXiv:1912.08772* .

Finamor, L. (2023). Labor market conditions and college graduation: evidence from brazil, *Economics of Education Review* **94**: 102403.

Fryer, R. G. et al. (2017). Management and student achievement: Evidence from a randomized field experiment, *Technical report*, National Bureau of Economic Research.

Grissom, J. A., Egalite, A. J., Lindsay, C. A. et al. (2021). How principals affect students and schools, *Wallace Foundation* **2**(1): 30–41.

Grissom, J. A. and Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills, *American Educational Research Journal* **48**(5): 1091–1123.

Henriques, R., Carvalho, M. d. and Barros, R. P. d. (2020). *Avaliação de impacto em educação: a experiência exitosa do programa Jovem de Futuro em parceria com o poder público*, Instituto Unibanco, São Paulo, SP.

Jackson, C. K. (2010). Do students benefit from attending better schools? evidence from rule-based student assignments in trinidad and tobago, *The Economic Journal* **120**(549): 1399–1429.

Jackson, C. K., Kiguel, S., Porter, S. C. and Easton, J. Q. (2024). Who benefits from attending effective high schools?, *Journal of Labor Economics* **42**(3): 717–751.

Lavy, V. (2020). Teachers' pay for performance in the long-run: The dynamic pattern of treatment effects on students' educational and labour market outcomes in adulthood, *The Review of Economic Studies* **87**(5): 2322–2355.

Machado, C. and Szerman, C. (2021). Centralized college admissions and student composition, *Economics of Education Review* **85**: 102184.

Moreira, D. (2019). Success spills over how awards affect winners' and peers' performance in brazil. Working paper.

Oliva, B. T. (2014). *Três ensaios em economia da educação*, PhD thesis, EESP-FGV, Brasil.

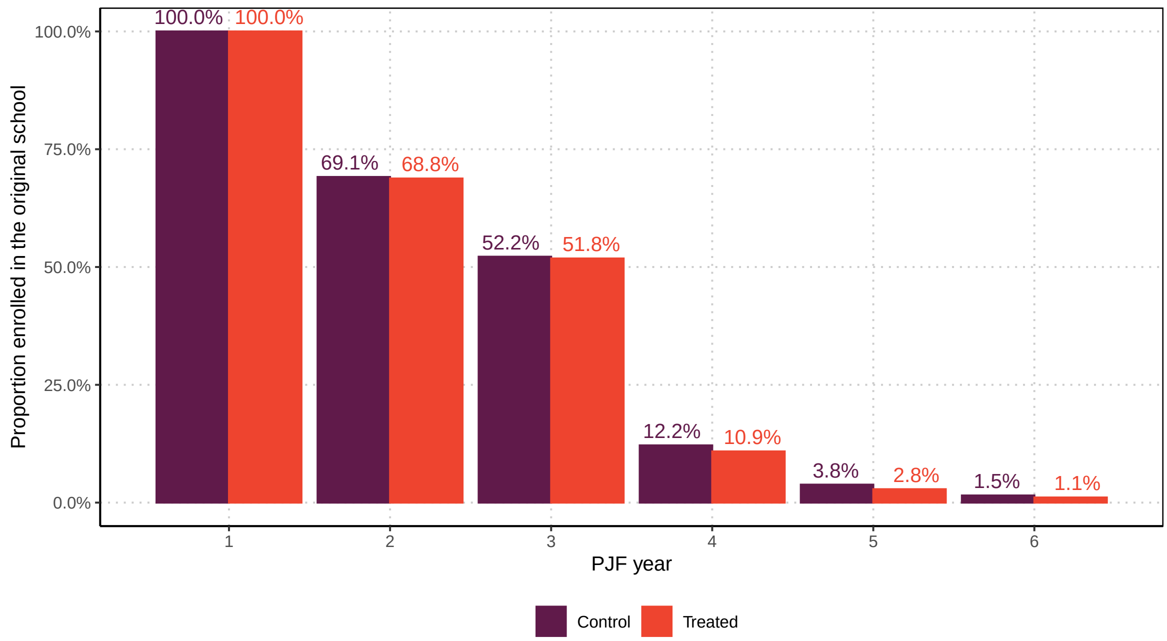
Rosa, L. S. N. (2015). *Avaliação de Impacto do Programa Jovens de Futuro*, Master's thesis, IPE-USP, Brasil.

Sass, T. R., Zimmer, R. W., Gill, B. P. and Booker, T. K. (2016). Charter high schools' effects on long-term attainment and earnings, *Journal of Policy Analysis and Management* **35**(3): 683–706.

Silva, S. (2010). *Análise dos efeitos de programas educacionais: o caso projeto Jovem de Futuro do Instituto Unibanco*, Master's thesis, EESP-FGV, Brasil.

Online Appendices

Figure A.1: Proportion of students enrolled in the original assigned school by year and type of school



Notes: Proportion of students enrolled in the original assigned school in the first year of high school by years and type of school (treated or control schools). This sample considers the entire sample, that is, including individuals that dropped out of school in the first year.

Table A.1: Robustness - Different specifications

Outcome	In Sample	Has CPF	HS Graduation (any year)		
	(1)	(2)	(3)	(4)	(5)
Treat	−0.007	−0.006	0.020	0.016	0.022
(s.e.)	(0.010)	(0.006)	(0.011)	(0.010)	(0.012)
[p-value]	[0.505]	[0.279]	[0.074]	[0.107]	[0.066]
N Obs	78,646	65,435	65,435	78,646	52,545
N Schools	207	207	207	207	207
N Strata	84	84	84	84	84
Control Mean	0.833	0.803	0.503	0.431	0.540
Sample	Full Sample	Study Sample	Study Sample	Full Sample	Study Sample with CPF

Notes: The table shows the estimates of 1, displaying the β coefficient, the standard errors (clustered at the strata level), and the corresponding p-value. Each column shows a different outcome and a different sample. In the first two columns, the outcome variable is an indicator for being in our main sample used in the study, and having a CPF. In the third, fourth, and fifth columns, the outcome variable is ever graduating from high school. In terms of sample in columns 1 and 4, the entire sample is used (all students enrolled in the first year of high school). In columns 2 and 3, the study sample is used, removing students who dropped out of high school in the first year. In the last column, we use the study sample for students for whom we have their CPFs.