Teacher-Led Innovations to Improve Education Outcomes: Experimental Evidence from Brazil☆

Caio Piza[*,a], Astrid Zwager[a], Matteo Ruzzante[b], Rafael Dantas[a], and Andre Loureiro[c]

[a]*Development Impact Evaluation (DIME), The World Bank, 1818 H Street NW, Washington, DC 20433, United States*
[b]*Northwestern University, Economics Department, 2211 Campus Drive, Evanston, Illinois 60208, United States*
[c]*Education Global Practice, The World Bank*

Revised April 2024

Abstract

We provide experimental evidence from an education program in Brazil that empowers public school teachers, through a combination of technical assistance and earmarked funding, to design and introduce locally adapted pedagogical innovations. While the study encompasses grades 5, 6, and 10, we find consistent and pronounced impacts on learning and school progression in 6th grade, a critical transition year from primary to lower-secondary education. Positive effects are concentrated in schools where teachers are most affected and where the rate of in-school project implementation was highest. We argue that program components are likely complementary and that education projects designed to tackle multiple constraints simultaneously can improve service delivery and child outcomes.

Keywords: education policy, lower-secondary education, autonomy, mentoring, school resources, teacher motivation, achievement, socio-emotional skills.

JEL Codes: H52, I21, M52, O15.

☆This draft benefited from comments from Pablo Acosta, Guadalupe Bedoya, Paul Christian, Flavio Cunha, Steven Glover, Florence Kondylis, Arianna Legovini, John Loeser, Pedro Olinto, Daniel Rogger, Ravi Somani, and seminar participants at DIME, São Paulo School of Economics, the 5th IZA Workshop: The Economics of Education, the XIII RIDGE Forum: 14th Annual Meeting Impact Evaluation Network of LACEA, Northwestern University, Universidade Federal de Pernambuco, and the 2022 Special ESA Meeting. We thank the World Bank i2i fund and the World Bank Brazil Country Office for generous research funding. Finally, we thank the staff at the Rio Grande do Norte State Secretariat of Education for being great research partners. The paper was first circulated under the title "Supporting Teacher Autonomy to Improve Education Outcomes: Experimental Evidence from Brazil". Computational reproducibility was verified by DIME Analytics. The reproducibility package is available on the GitHub repository `brazil-pip-education`. The trial was registered in the AEA RCT Registry under the unique identifying number AEARCTR-0007263. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

*Corresponding author.

E-mail addresses: caiopiza@worldbank.org (C. Piza); azwager@worldbank.org (A. Zwager); matteo.ruzzante@u.northwestern.edu (M. Ruzzante); rdantas@worldbank.org (R. Dantas) aloureiro@worldbank.org (A. Loureiro).

1 Introduction

Over the last three decades, many countries have succeeded in putting children in school yet learning gains have been limited (Angrist et al., 2021). Improving the quality and years of education is a priority for many countries given its role in building human capital, affecting individual earning prospects and long-term growth (Hanushek and Woessmann, 2008). Despite increasing resource allocation to education, governments have struggled to substantially improve education outcomes (McEwan, 2015; Glewwe and Muralidharan, 2016). The recent *World Development Report* points to both a "learning crisis" faced by many countries and the urgent need for solutions (World Bank, 2018).

Attempts to improve student outcomes often focus on increasing teacher effectiveness due to their central role in the education production function (Chetty et al., 2014; Araujo et al., 2016; Jackson, 2018; Bau and Das, 2020). This goal can be pursued by improving teacher skills and pedagogy and/or by providing (monetary or non-monetary) incentives to strengthen teacher motivation (World Bank, 2018). We present experimental evidence of an education policy in Brazil that provided teachers with support to autonomously design and implement a local project to tackle their specific issues instead of a centrally defined curriculum ("prescribing solutions"). The program encouraged teachers to propose pedagogical innovations relevant to their specific context, with the goal of improving both student progression and learning outcomes through increased teacher and student engagement.

However, the effect of giving local civil servants more incentives, such as autonomy, on the quality of service delivery is an empirical question. On the one hand, increasing the autonomy of local staff could lead agents to exert low effort due to the limited ability of the central government to observe and reward effort accordingly. For example, decentralization of the decision-making process may backfire if resources are captured by local entities or used inefficiently (Burgess et al., 2012; Banerjee et al., 2021). On the other hand, greater autonomy could improve service delivery by providing a non-monetary incentive for agents and adding meaning to the job (Cassar and Meier, 2018) or by leveraging their superior knowledge of local context (Duflo et al., 2018; Rogger and Somani, 2018).[1] Rasul and Rogger (2018) and Rasul et al. (2021) find that more autonomy is positively correlated with the quality and completion of public projects delivered even in contexts of low government capacity. While Bandiera et al. (2021) suggests autonomy can reduce the misalignment of incentives between officials and taxpayers with potential welfare benefits for society.

We study the Pedagogical Innovation Project (*Projeto de Inovação Pedagógica*—PIP), which encouraged teachers to propose pedagogical innovations relevant to their specific context, with the goal of strengthening student-teacher interactions. The program was envisaged combining teachers and pedagogy, two critical ingredients in the education production function (Banerjee et al., 2016).[2] The implementation of such a program is particularly interesting in the context of the Brazilian education system, in which most public schools continue to embrace the "traditional pedagogical method", characterized by a hierarchical relationship between teachers and students, with only a passive role for students, a centrally defined curriculum ("prescribed solutions"), and a limited degree of freedom for teachers to experiment with different pedagogical activities (Carvalho, 2016).

The program was implemented by the State Secretariat of Education (SEE) of Rio Grande do Norte (RN) in Brazil. RN consistently scores at the bottom of the Brazilian Education Development Index (*Índice de Desenvolvimento da Educação Básica*—IDEB).[3] While its cornerstone was to provide teachers with autonomy, the program was designed as a comprehensive package, including technical assistance and a limited grant, to support

[1]The association between autonomy and intrinsic motivation is at the foundation of Self-Determination Theory in the social psychology literature (Deci and Ryan, 1985). Seminal studies have focused on how monetary rewards might crowd out motivation, as they undermine autonomous decision-making (Deci, 1971), and how non-monetary incentives, which give greater autonomy, can enhance motivation (Zuckerman et al., 1978).

[2]Unlike remedial education programs, such as teaching at the right level, that tend to directly address reading and math skills, the PIP's pedagogical initiatives were designed to be entertaining and engaging as explained below.

[3]IDEB is a national indicator for the quality of education and combines information on student test scores and passing rates. Established in 2007, it has become one of the principal outcomes for the design of Brazilian educational policy, setting targets for schools, municipalities, and states.

teachers in the design and implementation of the proposed activities. Through seminars and support from a dedicated mentor, teachers developed a diagnostic of their main pedagogical challenges and context-specific innovations to address them. Schools were encouraged to introduce innovative pedagogical projects and expose students to non-curricular competencies and learning opportunities outside the classroom, such as the development of school radio and video production, the setup of theater plays or book fairs, and robotics classes. Mentors complemented local capacity while ensuring close ties with central government, possibly reducing moral hazard concerns associated with the strategic behavior of local staff. Approved proposals were awarded financial support to implement the projects, ranging from about 7,500 to 11,000 US dollars (USD),[4] or median USD 139 per student, i.e., 3.6 percent of average annual expenditure per student in Brazil (OECD, 2016).

Our experiment focuses on the 2016 iteration of the program, which targeted the final grade of primary education (5th grade), the first grade of lower-secondary education (6th grade), and the first grade of upper-secondary education (10th grade), with the latter two generally being the most problematic in terms of repetition and dropout rates, according to the school census (INEP, 2015). Of 299 schools eligible for the program in 2016, 130 schools were randomly invited to participate and submit a proposal. Schools were included in the selection with the highest participating grade they offered, and the pedagogical projects were only implemented in that grade. Randomization was then stratified by grade.

We first show that the program positively impacted student learning and school progression, the outcomes targeted by the program. To assess the program's effects on student learning, we use the state's standardized exam that was introduced in 2016 and extended to grades selected for the impact evaluation. The intent-to-treat (ITT) estimates show no overall impacts, but splitting the analysis by grade reveals substantial impacts in grade 6: a 0.18 SD improvement to math scores and a 0.16 SD improvement to Portuguese scores.

[4]Equivalent to 30,000 to 45,000 Brazilian *reais* (BRL), using the exchange rate on December 31, 2015.

Slightly lower impacts are observed on humanities (0.10 SD) and natural sciences (0.12 SD). We estimate the average impact on learning to be equivalent to half an extra year of schooling, or 0.36 years per USD 100 spent.

Overall passing rates increased by 4.70 percentage points (pp), a 6.6 percent improvement over the control mean of 71 percent. When disaggregating the results, we find these are also driven by improvements in grade 6, with passing estimated to increase by 8.46 pp, a 13 percent improvement compared to the control mean of 63.6 percent. A back-of-the-envelop calculation of the combined effect of increased learning and higher probability of finishing high school for 6th graders suggests a net present value (NPV) of the expected years of schooling on future earnings ranging between USD 7–13,000, or 28 to 52 times the annual Brazilian minimum wage. Compared to the cost of the program per student (USD 139), the estimated NPV suggests that PIP was a high-return investment for the state.

We try to unpack which program components and mechanisms might be driving these results and their concentration in grade 6. First, we hypothesize that the program increases teacher retention if teachers feel more committed to implementing their own pedagogical projects during the academic year. Pooling all grades, the ITT estimate on teacher retention is positive but not statistically significant. However, in line with learning and progression results, we find a 15.5 percent increase in teacher retention for grade 6, which is driven by schools with low teacher retention at baseline. A similar, though less precisely estimated, pattern is observed for grade 10.

Second, using administrative data, we find that in-school project implementation was higher in 6th-grade schools yet is particularly poor in 10th-grade schools with low teacher retention at baseline.

Third, we leverage the fact that most 6th-grade teachers also teach other grades. Mechanically, we find a similar increase in 7th-grade teacher retention, yet do not find any (positive or negative) spillover impacts on the progression rates of 7th-grade students.

This provides suggestive evidence that increased teacher engagement may not necessarily spillover to other grades in the absence of the in-school developed projects.

Taken together, our findings indicate that a teacher-led approach can work, particularly if complemented with other inputs; the targeted scope of the projects and technical assistance to support operationalization might be instrumental for success. Conversely, lack of school budget is unlikely to be a key driver of our main results given that most treated schools implemented the projects with existing resources during the school year due to substantial delays in the transfer of the allocated grant and they typically struggle to disburse all the federal funding they have available. However, the potential of the operational support is highlighted by the fact that treatment schools were able to overcome general administrative hurdles. In fact, we find substantially higher general funds disbursement in the year following the program.

Given the program's emphasis on changing student-teacher interactions through innovative pedagogical approaches, we also test whether the program impacted students' socio-emotional skills, as these could be either directly impacted through improvements to both teacher-student interactions and students' motivation or indirectly impacted through changes to cognitive skills (Cunha and Heckman, 2007). To do so, we measure the Big Five personality traits. Pooling all grades, we find that the program had a positive effect on conscientiousness and extroversion. For grade 6, we find conscientiousness increased by 0.17 SD., the trait most commonly associated with the acquisition of cognitive skills (Poropat, 2009; Ivcevic and Brackett, 2014), and extroversion increase by 0.20 SD. Our results indicate that the intervention was mostly successful for students in grade 6, a critical grade for students as they transition from primary to lower-secondary education when students move from having a single teacher to multiple teachers (Bedard and Do, 2005; Hanewald, 2013; Santos et al., 2017). Improving teacher and student motivation might therefore counterbalance the weakening of student-teacher interaction at this stage.

Our results indicate that efficiency gains in education delivery can be obtained by leveraging mostly existing systems and resources. We show that combining the autonomy of civil service providers with targeted technical assistance and funds can improve outcomes of interest even in a low-capacity environment.[5,6] Recent attempts to improve traditional teaching practices predominantly relied on training and structured approaches, including remedial education programs (e.g., Banerjee et al., 2007; Banerjee et al., 2017; Marinelli et al., 2021), technology-aided instruction (Muralidharan et al., 2019; Beg et al., 2022), and standardized lesson scripts (Gray-Lobe et al., 2022). The pedagogical interventions we study harness teacher-led innovations, which are customized to address school-specific issues. Complementing local capacity on how to design and operationalize pedagogical projects may be critical, as teachers' autonomy alone has had limited success in the Brazilian context (Almeida et al., 2016; Oliveira et al., 2016).

Second, our paper highlights the importance of thinking about bundled educational policies that accommodate the multi-faceted constraints students face in their specific context and provide comprehensive support to local implementers. Our exploration of the mechanisms reveals that the distinct components of the program likely complemented each other in achieving positive impacts on cognitive and socio-emotional learning in 6th grade. This aligns with growing evidence on the importance of complementarities between school inputs and teacher incentives in education production (Mbiti et al., 2019, Gilligan et al., 2022).

Finally, while previous experimental evaluations of school grants have shown mixed results, such programs hide a wide degree of heterogeneity in terms of grant size, design features, and decision-making responsibilities within the school.[7] The key novelty of the intervention examined in this paper is that the allocation of funds was bound to pedagogical activities designed and implemented by teachers as opposed to school management or

[5]For example, disbursement rates of federal funds allocated for investments in school infrastructure and pedagogy are, on average, as low as 30.5 percent in this context.

[6]The paper does not speak to the wide literature on school decentralization, which involves allowing local management of resources and/or curriculum. Existing studies on autonomy in the public sector are reviewed above.

[7]Glewwe et al. (2009) in Kenya, Das et al. (2013) in India (for anticipated grants), Blimpo et al. (2015) in The Gambia, Beasley and Huillery (2017) in Niger, and Mbiti et al. (2019) in Tanzania found null results on student learning, as measured by test scores. In contrast, Das et al. (2013) in India (for unanticipated grants), Carneiro et al. (2020) in Senegal, and Andrabi et al. (2021) in Pakistan reported promising improvements in student outcomes.

specific items mandated by the central government. The one-time grant was conditional, i.e., the money could not be spent on other school expenses, such as teacher salaries or infrastructure improvements.

The remainder of the paper is organized as follows. Section 2 details the context and intervention. Section 3 describes the experimental design and data sources. Section 4 presents the empirical strategy and main results, while Section 5 explores potential mechanisms driving the main findings. Section 6 provides back-of-the-envelope estimates for the impact of the program on school quality indicators and individuals' expected earnings. Lastly, Section 7 concludes with policy recommendations.

2 Context and Intervention

2.1 Education in Brazil and Rio Grande do Norte

While Brazil has made significant strides to guarantee universal access to primary education, reaching a 99 percent enrollment rate for children aged 6–14 in 2018 (IGBE, 2018), substantial challenges remain to keep children in school and ensure the quality of education. Grade repetition and dropout rates in primary and secondary schools are among the highest in the Latin America and the Caribbean (LAC) region (UNESCO, 2020). Despite the largest improvements in math scores in the Program for International Student Assessment (PISA) between 2003 and 2012, Brazil still ranks below all LAC countries except for Peru and the Dominican Republic (OECD, 2015).

These national figures hide a high degree of regional variation. In this paper, we study an education program implemented by the RN state government, one of Brazil's poorest states. In the 2015 national standardized exam,[8] RN state schools scored at the bottom of the learning distribution in both primary and lower-secondary education.[9] The difference in 5th-grade proficiency levels between the average student in RN and the best-performing

[8] *Sistema de Avaliação da Educação Básica* (SAEB).

[9] 2015 is the year prior to the roll-out of the interventions we study in this paper.

state is the equivalent of 2.5 years of education.[10] The low level of learning is reflected in the state's progression indicators. In 2015, the average school dropout rate in upper-secondary education was 12.4 percent compared to the national average of 8.8 percent (INEP, 2015). The combination of high dropout rates and low learning outcomes puts RN state schools near the bottom of the Brazilian Index of Development of Basic Education (online Appendix Figure B1).

Grade repetition is particularly high in grade 6, peaking at nearly 32 percent (online Appendix Figure B2a). Grade 6 marks the first year of transition from primary to secondary school. Children move from having one dedicated teacher to having one per subject. On the other hand, grade 10, which is the first year in upper-secondary (high school), is when most dropouts occurs (online Appendix Figure B2b).

A major constraint to school quality and student achievement in Brazil is principal and teacher turnover, which is around 21 percent nationally (Akhtari et al., 2022). In the RN public school system, 30 percent of teachers leave their schools each year, with little variation across grades (online Appendix Figure B3), potentially disrupting school operations and compromising personnel collaboration.[11] Using school-level data from INEP, we find that teacher permanence is positively correlated with student passing rates and negatively correlated with age-grade distortion, repetition, and dropout, for both primary and secondary schools (online Appendix Table D1).[12]

2.2 The Pedagogical Innovation Project (PIP)

The Pedagogical Innovation Project (*Projeto de Inovação Pedagógica*—PIP), developed by the RN SEE, aimed at improving both student progression and learning outcomes by increasing child and teacher engagement. The intervention has four main components:

[10]This uses the calculation proposed by Alves et al. (2016).

[11]One reason for the high turnover relates to how the placement of teachers is organized in Brazil. Teachers are initially placed at any school with a vacancy, with limited consideration of their location preferences. Then, every year, teachers are allowed to compete for new vacancies.

[12]Teacher permanence is an index produced by INEP. It averages, at the school level, the number of years a teacher stays in a given school over a five-period period, weighting for the number of teachers in a school. The index ranges from zero to five, where a higher number indicates more regularity of the teacher pool in a school.

(i) a high degree of autonomy for teachers to design and implement a project based on their diagnostic of the context-specific challenges; (ii) the introduction of pedagogical innovations in the classroom; (iii) continuous technical support to teachers during the design and implementation of the project, with the SEE having only an advisory role to assure minimum quality standards; and (iv) a grant specifically earmarked to implement the project.

The approach of PIP sought to ensure the relevance of the interventions and motivate teachers and students. The program design is based on the premise that: (i) school staff are better equipped than central-level bureaucrats to identify solutions to school-specific problems using local knowledge; (ii) allowing autonomy over the selection and development of interventions motivates teachers by giving them the opportunity to implement activities that leverage their local knowledge; and (iii) innovative projects can engage students and improve student-teacher interactions.

PIP targets primary and secondary state schools—16 percent of primary schools, 41 percent of lower-secondary schools, and 94 percent of upper-secondary schools in the public education system. The program has been implemented in grades 4, 5, 6, and 10, the grades with the most critical dropout and repetition rates. PIP was launched in 2014 and between the 2015 to 2018 school years covered 397 of the 639 state schools.

The SEE supported teachers during project development and implementation. Here we detail the support in each of these phases.

2.2.1 Project Development

To initiate the design phase, schools are invited to participate in a three-day workshop on innovative and project-oriented teaching practices. During break-out sessions, participants identify the main pedagogical challenges they face and discuss how the innovation concepts would fit their context. Each school is provided with an individualized report card comparing its test scores and passing grades with the average of the state, region, and city.

Following the workshop, each school is assigned a mentor (*professor orientador*) to support the development of the innovative project. The mentors are part of the SEE central team and each is assigned to 10 schools on average.[13] First, teachers prepare a diagnostic of their challenges, such as low academic performance, grade repetition, indiscipline, lack of motivation, or school dropout. Based on the diagnostic, teachers identify possible drivers and propose an innovative and actionable plan to improve the targeted education outcomes. The mentor then works with the school to translate the diagnostic and proposed project into a detailed implementation plan that is reviewed by the SEE of RN.

2.2.2 Implementation Support and Monitoring

Schools with approved proposals are awarded a fixed amount of funding to execute their projects. Schools can only spend the operating funds on inputs directly related to their proposed project. The grant amount depends on the number of classes included in the project and ranges from BRL 30,000 to 45,000, i.e., USD 7,576 to 11,364 (online Appendix Figure B4). The median transfer per enrolled student was BRL 555.55, the equivalent of USD 139, which represents about 3.6 percent of average annual expenditure per student in Brazil (OECD, 2016).

Through subsequent visits and remote follow-up, mentors closely support the implementation of the projects. Mentors help schools obtain the necessary paperwork to access the funding and prepare procurement of materials.

2.2.3 Characteristics of Sub-Projects

Schools were encouraged to explore teaching settings beyond traditional lecture-style lessons to improve student-teacher interactions and to embed their project across disciplines, increasing coordination across subjects. Proposed projects were evaluated by the SEE. The project had to demonstrate an innovative methodology for that school's context, and not necessarily a frontier methodology. All submitted proposals were approved.

[13] Mentors are selected based on their experience with implementing pedagogical projects in schools and all are existing staff of the state secretariat.

Most proposals fell into one of the following three categories:

· *Writing and reading*: These sub-projects were designed to improve students' literacy and oral communication skills. They included activities such as studying Brazilian literature classics, publishing school newspapers, broadcasting a school radio, setting up theater plays, or organizing book fairs and poetry contests.

· *Communication, media, and culture*: The focus of this type of sub-project was to introduce students to modern-day digital tools and give teachers the opportunity to use new technologies and social media. Examples include developing video games and robotics classes.

· *Culture and arts*: The goal of these sub-projects was to explore different forms of cultural and artistic expressions, such as painting, graffiti, dance, theater, cinema, and music. Examples of pedagogical projects are detailed in online Appendix A.

Project activities were usually integrated into normal school hours to not create any supplemental time burden for teachers. The treatment dosage was also the autonomous decision of the teachers.

3 Experimental Design and Data

The PIP was first launched in 2014 with implementation taking place in the 2015 school year. Each year, a subset of state schools were invited to join the program. Our study focuses on the cohort of schools that were eligible to initiate design in 2015 for project implementation in the 2016 school year. That year, only grades 5, 6, and 10 were included. This section further details the selection of participating schools and data sources.

3.1 Experimental Design

To ensure enough operational capacity, only a sub-sample of schools was selected to participate each year. To determine the pool of eligible schools for implementation during the 2016 school year, three filters were applied. First, only schools that would not change principals between the 2015 and 2016 school year were included to ensure continued buyin for the prepared projects. State legislation requires directors to change schools every two years, resulting in about half the schools changing directors each year.[14] Second, the 2016 edition targeted the final grade of primary education (5th grade), the first grade of lower-secondary education (6th grade), and the first grade of upper-secondary education (10th grade).[15] Only schools offering at least one of those three grades were considered. Finally, schools that participate in the Federal program ProEMI (*Ensino Médio Inovador)* were excluded.[16] As a result, of the 639 state schools, 299 were eligible to receive the PIP program in 2016.

The final selection of participating schools was done randomly among eligible schools, which forms the basis of our identification strategy. The RN SEE aimed to support a total of 130 schools in the 2016 school year. The randomization was stratified by school grade and region. From the 2015 PIP cohort, we learned that schools participate in just one grade. The SEE preferred to focus on higher grades, which is typically where schools experience more challenges. Therefore, schools offering several of the target grades (5, 6, and 10) are included in the randomization only with the highest target grade they offer.[17]

The state is divided into four regions and, combined with the three grade levels, this resulted in a total of twelve strata. In each stratum, around 40 percent of the schools were allocated to the treatment group. Larger schools may have more than one class in a grade, in which case all classes, and thus students, in the selected grade participated. Not all teachers of a grade necessarily participated. The selection of teachers to include in the program is decided within schools and is unlikely random. When analyzing student and teacher outcomes we always consider all students and all teachers of the selected grade.

[14]Mechanically, none of the schools from the first 2015 cohort were considered, since those were not change between the 2014 and 2015 school years but would between the 2015 and 2016 school years. This legislation has since slightly changed to allow for directors to stay on longer.

[15]Other editions of the program included 4th grade.

[16]*Ensino Médio Inovador* (Innovative High School project – ProEMI) was established in 2009 by the Ministry of Education as a policy aimed to support innovative curricular projects in upper-secondary schools through technical and financial assistance.

[17]For example, schools offering both grades 6 and 10 are only included with their grade 10 to the relevant stratum for randomization. Grade 6 in this case does not participate in the program and is not considered in the evaluation sample.

The randomization resulted in 130 eligible schools in the treatment group and 169 in the control group (Panel A in online Appendix Table B1). All 130 selected schools were invited to the workshops held in the final months of the 2015 school year. The randomization was performed using the 2015 school census. After the start of the 2016 school year, a few schools had closed or no longer offered the grade that had been selected for the intervention.[18] This leaves us with a final sample of 280 schools effectively allocated to the experiment at the beginning of the 2016 school year (Panel B in online Appendix Table B1)—126 in the treatment group and 154 in the control group. The geographical distribution and treatment assignment of these schools are shown in the online Appendix Figure B5. Across the selected grades in each school, 19,899 students were included in the experiment—9,432 in treated schools and 10,467 in control schools (Panel C in online Appendix Table B1).

3.2 Data

To assess the impact of the PIP, we leverage three main sources of data. We use administrative data, such as the Brazilian school census and information from the SEE, and collect data on cognitive and socio-emotional skills.

Administrative Data. We use administrative data from both the state's education monitoring system and the annual national school census to obtain school, teacher, and student characteristics and progression. The state's monitoring system, the *Sistema Integrado de Gestão da Educação* (SIGEduc) portal, provides data on passing, dropout, and repetition rates at the grade level.[19] The school census is carried out on an annual basis by the *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) of the Brazilian Ministry of Education. It contains information on overall school characteristics, such as location, presence of a library, science lab, and internet, as well

[18] Eight schools had closed, six were not offering regular classes anymore, four were selected for the 5th-grade experimental group but were not offering 5th grade anymore, and one was in the 6th-grade group but was not offering 6th grade anymore.

[19] Progression rates are reported at the end of the school year (i.e., February-March) by principals, and then validated by INEP.

as information on the number of teachers, students, and classes.[20] The census also allows us to track individual teachers and students over time, even if they move to other schools within the state.[21] Where possible, the analysis of the results uses both sources.

Using the census data, we define "teacher retention" for a given year as a dummy of whether a teacher is in the same school in the following year. The dummy is one if a teacher is still teaching in the same school (in any grade) and zero otherwise.

Finally, the SEE provided data on school directors and on the implementation of the PIP, such as the proposal score, resources allocated to schools, and execution of the projects. The rate of implementation of the proposed plan is assessed by the mentor at each visit.

Learning Outcomes. To measure student learning, we use the state standardized external assessment in math, Portuguese, human sciences, and natural sciences. The standardized exam was introduced at the state level in 2016. While it typically includes 5th and 10th grades, the exam was expanded to 6th grade for the purpose of the evaluation. It is administered yearly in October/November, i.e., at the end of the school year, and comparable to the national standardized test (SAEB).[22] All students are encouraged to take the exam as the school (and the municipality) only has its average grade considered for state-level reporting if at least 80 percent of students complete the assessment. However, the exam is not high-stakes for the students. We discuss potential issues related to student participation in online Appendix C. The test is scored on a 0–400 scale. We standardize the test score (within grade) so that the regression coefficients can be interpreted in terms of standard deviations from the control group.

In addition to the standardized measure of learning, we obtained students' final-year GPAs from *SIGEduc*. These are cross-subject scores, as evaluated by their teachers, on

[20]We extract school location and distance from the state's capital, Natal, by scraping the Google Maps API with school names.

[21]The Brazilian Education Census is implemented in two stages. At the beginning of the school year (i.e., May–July) initial student enrollment data are collected, and the survey of school, teacher, and students' characteristics is administered. In February–March of the following year, data are collected on passing/retention and on "movement", which includes dropouts and transfers.

[22]For math and Portuguese, we obtain the scores rescaled to SAEB, which allows us to put the impact on student learning in the Brazilian-wide context. Sciences are not included in the national exam.

a 10-point scale. Importantly for the attrition analysis featured in the next sub-section, this variable is reported for each single student in our experimental sample and is not restricted to the sample who remain in schools until the end of the year and so participates in the standardized assessment.

Socio-Emotional Skills. To analyze the impact on socio-emotional learning, we measure the Big Five personality traits (neuroticism, extroversion, conscientiousness, agreeableness, and openness). We use an off-the-shelf self-reported test developed and adapted to younger students in Brazil by the *Instituto Ayrton Senna*.[23] This test, and its equivalent, are widely used in the literature to assess socio-emotional skills.[24,25] The test was administered at the end of the 2016 school year to the grade that entered the randomization (see Section 3.1). In case a school had multiple classes in the same grade, one class was randomly chosen for the test.

3.3 Validity of the Experiment

Balance. To examine whether the randomization resulted in balanced samples across control and treatment groups, we compare observable characteristics prior to the roll-out of the program. Online Appendix Table C1 shows several characteristics at the school, grade, teacher, and student levels, including some of the key outcomes of the intervention, such as repetition and dropout rates. For grade, teacher, and student comparisons, we only consider the classes in the eligible grade for that school (see description in Section 3.1). Columns (2) and (4) show the means in the treatment and control groups. In column (5), we report both standard p-values based on t-test of differences in the means and p-values computed using randomization inference. Generally, we find no statistical

[23] The test is generic and does not reference PIP-related or promoted activities.

[24] See Kautz et al. (2014) for a review of the recent advances in measuring socio-emotional skills.

[25] Research has shown that individuals with the same level of a trait may assess themselves at very different levels on a Likert scale (Primi et al., 2016). To address this issue, we administered a set of anchoring vignettes that help reveal the respondent's latent scale and response style, allowing us to calibrate individual responses following the method suggested in Primi et al. (2016). The vignettes describe three hypothetical individuals that represent three distinct points on a scale (low, medium, and high). Students are asked to assess the personality trait of each of the characters along a 1–5 Likert scale. The student self-evaluation is then calibrated to a 1–7 scale according to her response to the vignette. In the analysis, we standardize these indicators (within grade): the resulting coefficients can be interpreted in terms of standard deviations from the control group.

differences when comparing the treatment and control groups. A joint significance test of school and student characteristics confirms that these variables do not jointly predict treatment assignment (F-stat of 0.69 and 1.76, respectively).

Randomization was done by grade level: to test the validity of the sub-group analysis, we also report p-values for the comparison in each grade in columns (6)–(8).[26] We find a statistically significant, yet small, difference in the age of 6th graders. The control group is, on average, 0.25 years older than the treatment group. In the analysis, we check the robustness of the results to the inclusion of this unbalanced variable as a control.

Compliance with Treatment Assignment. All 130 initially selected schools were invited to participate in the workshop, which occurred in late 2015. Of the 128 schools that attended, all prepared and submitted a proposal. All submitted proposals were approved, some after modifications. At the beginning of the 2016 school year, which starts in January, four of the 130 selected treatment schools had closed or did not offer the target grade anymore, resulting in a final sample of 126 schools, all with approved projects. Following approval, all schools received the first mentor visit at the beginning of the school year. Throughout the year, schools were meant to receive quarterly visits. Of the 126 schools, 109 received at least three visits during the school year, and 39 received all four visits. To receive the allocated funding, the schools had to provide proof that they did not have outstanding balances with federal, state, or municipal tax collection agencies.[27] The lack of this documentation delayed the transfer of operating funds for most schools. Transfers were supposed to occur toward the beginning of the school year in February, but the first transfers were only made in July. By the end of the 2016 school year, 90 schools had received the funding.[28] Despite the challenges with the transfer of resources, mentors worked with the schools to continue the implementation of the activities proposed in their work plan. By the end of the school year, 74.6 percent of schools had completely

[26] Firpo et al. (2020) show that, in stratified experiments, balance tests based on fixed effects regressions may not be sufficient to detect relevant imbalances because of lack of power. In that case, it is preferable to run balance checks at the stratum level.

[27] Although public schools do not pay taxes, they do need to file that they are exempt.

[28] Eight schools received the funding in the following year.

implemented the planned activities. All analysis takes into consideration the original assignment in the experiment and should therefore be interpreted as ITT effects.

Missing Data. Not all schools and students participated in the socio-emotional and proficiency test: 94 percent of schools in the evaluation sample participated in the state standardized tests and 84 percent in the socio-emotional test; among the participating schools, on average, 69 percent of enrolled students took the proficiency tests and 55 percent the socio-emotional test. We discuss attrition in test-taking and the resulting missing data on student-level outcomes in detail in online Appendix C and summarize the results and robustness checks in Section 4. We also explore whether changes in class composition as a result of the intervention, for example by reducing drop-out rates, might drive the results.

4 Empirical Strategy and Results

4.1 Empirical Strategy

We estimate the effect of randomly assigning schools to the intervention on our outcomes of interest with the following reduced-form specification,

$$y_{isb} = \alpha + \beta \cdot T_{sb} + \Sigma_b + \varepsilon_{isb} \quad (1)$$

where y is the outcome of interest for student i in school s and strata (or block) b, T_{sb} is the indicator variable of treatment assignment, Σ_b is a vector of strata dummies, and ε_{isb} is the error term. Standard errors are clustered at the school level, the level of randomization.[29]

There are three potential sources of non-compliance with the treatment. First, not all assigned schools received all components of the program, as discussed in Section 3.1. Second, not all teachers at an assigned school participated in the program. Finally, while participating teachers were supposed to apply the pedagogical projects in all their classes, there might be variation in treatment exposure/intensity due to variation in the

[29]Some estimates are obtained at the school level. In these cases, we employ robust heteroskedasticity-consistent (Eicker-Huber-White) standard errors.

number and types of teachers participating. To account for this, we include all schools, teachers, and students of the assigned grade, per their original assignment. Therefore, the parameter β identifies the ITT effect. To estimate treatment-on-the-treated (TOT) effects, we use the school random assignment as an instrumental variable to account for non-random variation in teacher participation in the program.

We provide estimates of program impact for all schools pooled as well as for each grade separately. For grade-specific estimations, we present p-values corrected for multiple hypothesis testing across grades, following the step-down procedure described in Romano and Wolf (2005, 2016). To examine sensitivity to imbalance in missing learning and socio-emotional data, we bound treatment effects by adjusting for differential attrition, as proposed by Lee (2009), and we estimate confidence intervals around such bounds, which capture both uncertainties about potential selection bias from missing data and sampling error, following Imbens and Manski (2004). To check the robustness of the results, we then estimate the model by adding controls,[30] and we use blocked difference-in-means, interaction-weighted and regression-weighted estimators (IWE and RWE, respectively).[31]

To explore the potential distributional effects of the program, we estimate unconditional quantile treatment effects (UQTE) following Firpo et al. (2009). Unlike the average effect, quantile treatment effects assess whether the impact of the program differs at distinct points (quantiles) of the outcome distribution. The UQTE has a similar interpretation as the average effect and is estimated by computing the horizontal difference between accumulated (or marginal) distributions of treated and control outcomes for a given quantile.

[30]The covariates included are student's age, gender, and race dummies (white, indigenous, black, or *pardo*), whether they receive *Bolsa Família*, and whether they use school transportation.

[31]The blocked difference-in-means approach uses strata sizes, instead of fixed effects, to weight the treatment effects estimates within each stratum. Gibbons et al. (2018) show that, in the presence of heterogeneous treatment effects, fixed effects estimates are generally not a consistent estimator of the average treatment effect. Therefore, they propose IWE and RWE as alternatives to recover such parameter.

4.2 Results

We first present the estimates on the main student outcomes the program targeted, student learning and progression. We then explore which mechanisms might have contributed to these results.

4.2.1 Learning Outcomes

Table 1 shows ITT estimates on overall test scores, by subject and by grade-subject. We find a large, positive impact on learning outcomes, but for 6th graders only. The intervention improved overall test scores for 6th graders by 0.15 SD, or six points compared to the control mean of 163. In the next section, we describe results on student progression and explore whether changes in the composition of test-takers might be affecting the estimates. The coefficients for 6th graders survive multiple hypothesis corrections. For robustness, we re-estimate the model controlling for a vector of student covariates and using alternative estimation strategies, such as blocked difference-in-means, IWE, and RWE. The results are very similar and are available in online Appendix E.

Distributional impact analysis suggests significant gains are made across the board with a more pronounced impact at the higher end of the test score distribution (online Appendix Figures D1–D2). On average, the intervention positively affected learning outcomes of both female and male 6th graders (online Appendix Figure D3). However, the estimates suggest that the program shifted the entire distribution of boys' test scores to the right, but for girls, it resulted only in differences in the higher quantiles. The quantile estimates indicate that the program helped boys catch up with the initially higher proficiency level of girls.

To contextualize the magnitude of the impact on 6th graders, we convert the learning gains from the program into additional years of schooling. To do so, we use the state standardized test scores rescaled to the national standardized exams (SAEB). The exam is taken in grades 5 and 9 and is constructed to allow for the comparison of levels on a unique proficiency scale across grades and years.[32] This enables the calculation of the accumulated knowledge in math and Portuguese of an average student between the tests taken in 5th and 9th grades. To calculate the average gains in knowledge between those four years of schooling, we compare the test scores of a cohort of students from RN that took the 5th-grade exam in 2013 and the 9th-grade exam in 2017. We find that the average gain in test score for this cohort was 60 points, 15 points per year on average. Based on the ITT estimates, we find that PIP improved 6th graders' math and Portuguese scores by 6.83 and 6.78 points, respectively, on the SAEB exams scale, the equivalent of a little under half a year of additional schooling.[33] In Section 6, we reflect on the economic implications of these results.

4.2.2 Student Progression

The positive results on learning outcomes across the distribution of scores suggest that more students now pass the threshold to progress to the next grade. In combination with potentially direct impacts on motivation and engagement, this is expected to improve student progression rates.

To test whether the program affected student progression, we estimate impacts on grade passing, repetition, and dropout. Columns (1) and (5) of Table 2 show the ITT effects across grades. We report results using both data from SIGEduc, which are reported at the grade level (column 1), and from tracking individual students using the 2016 and 2017 waves of the school census (column 5). We find positive impacts on overall progression. Passing rates are estimated to increase by 4.70 pp, a 6.6 percent improvement over the control group mean, using SIGEduc data, and by 4.51 pp (7.5 percent) using census data. The impacts on grade passing mechanically result from either a reduction in dropout or

[32] The exam uses item response theory (IRT) to express scores on a unique scale for all grades of the national education system. This is achieved by including test items from 5th-grade tests into 9th-grade tests. The same is done from one edition to the next, making SAEB scores comparable over time. The test takes place every two years.

[33] Results using SAEB-rescaled test scores as the outcome variable are presented in online Appendix Table D2. In our data, one SD improvement in learning in 5th grade corresponds to 50 points, i.e., 3.3 years of schooling. Comparing gains in literacy for a set of countries, Evans and Yuan (2019) find that a one-SD improvement in test scores ranges from 4.7 to 6.5 years of schooling.

repetition or a combination of both. We find that the general results are driven by a reduction in repetition.

When disaggregating the result we find that the overall results in passing are largely due to substantial improvements in 6th grade passing rates, which are estimated to have increased by 8.46 pp, a 13 percent improvement compared to the control mean of 63.56 percent. The results using the census data are similar: a 7 pp increase among 6th graders (12 percent). The estimates pass multiple hypothesis testing, using either data source.

The SIGEduc data suggest that the 6th-grade result was mainly achieved by reducing grade repetition, while census data point to a reduction in dropout being the main driver. The discrepancy in the results can be explained by the difference in the timing of defining a student's status—the SIGEduc data only captures students dropping out during the school year, while tracking students into the next census wave also captures dropouts of students over the summer break. This suggests that some of the students reported as retained in SIGEduc drop out by the beginning of the next school year. While we do not observe significant impacts on overall passing in grade 10, census data point toward a reduction in repetition.

The same robustness checks used for estimating the impact on student learning can be found in online Appendix E and do not alter our findings. Further, we find no evidence of differential impacts by gender (online Appendix Table D3) or of heterogeneous effects by baseline levels of passing rate (online Appendix Table D4), suggesting that the score-cards distributed during the design workshop containing information on schools' relative performance (see Section 2.2.1) are not influencing the results.

The reduction in 6th-grade repetition might have long-term implications for students' years of education and likelihood of completing school. To explore how much improving progression may affect students' school careers, we track all RN students who were in 6th grade in 2011 up to 2017 using school census data. We find that students who were promoted in 6th grade in 2011 are 40 pp more likely to be in school in 2017 than students who were retained in 2011 (online Appendix Figure B6a). Similarly, after six years, they have completed 2.34 more years of schooling (online Appendix Figure B6b). We quantify the correlation between retention in 6th grade and schooling outcomes by estimating an OLS regression of dropout and completed years of schooling on grade repetition.[34] We find that failing 6th grade is associated with a 21 pp higher likelihood of school dropout after six years, and a reduction of 1.7 years of completed schooling (online Appendix Table D5). Taken at face value, our estimates provide suggestive evidence that the reduction of 23 percent (or 7 pp) in repetition rate caused by the PIP might contribute to substantially reducing school dropout (by 4.83 pp) and increasing years of schooling (by 0.4 extra years) of the treated cohort of 6th graders.

4.2.3 Robustness to Attrition and Changes in Sample Composition

The results on test scores we have described so far may be driven by a combination of actual improvements to cognitive skills or changes in sample composition as a result of the impacts on progression or selective attrition into testing. To distinguish between these interpretations, several empirical tests are presented in online Appendix C. The results suggest that attrition is likely not driving our findings.

First, while observing that schools and students participating in the test are a selected sample of the study population, we do not find any evidence of either differential or selective attrition between treatment and control schools. The share of schools and students participating in proficiency tests is balanced across treatment and control, and balance in baseline school and student characteristics is maintained in the sub-sample of test-takers. The pattern of selection into treatment is the same among treatment and control groups: participating schools and students appear to have better education outcomes at baseline, but such selection patterns are the same across experimental arms. Therefore, our results are likely unbiased estimates of program impacts among tested schools, yet they may not

[34] We estimate the following cross-section regression: $y_{isc} = \alpha + \beta \cdot retained_{isc} + \sigma_s + \gamma_c + \varepsilon_{isc}$, where y_{isc} is the outcome variable, i.e., dropout dummy or years of completed schooling, of student i in school s and class c, $retained_{isc}$ is a dummy variable for students who repeated 6th grade in 2011; σ_s and γ_c are school and class fixed effects. Standard errors are clustered at the school level.

extend to the non-tested schools.

Second, we find no program impact on dropout during the school year, limiting the likelihood of changes in the sample due to drop-outs not taking the exam in the control group. We leverage self-reported GPA and observe that students who drop out or repeat the grade are typically those with the lowest test performance in their class.[35] Hence, if anything, the program's impact on reducing drop-outs in 6th grade should reduce learning outcomes in treated schools and bias our estimates on standardized test scores downwards.

Finally, the treatment effects are robust to attrition: bounding the ITT estimates to account for potential bias due to missing test score data slightly widens the confidence intervals but does not substantially affect the significance level of the results (online Appendix Table C6).

[35] This pattern is observed in each grade.

Table 1: Impact on Student Learning

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
All schools					
Treatment	0.032	0.041	0.028	0.012	0.044
	(0.044)	(0.051)	(0.057)	(0.039)	(0.039)
Number of observations	12760	11366	11365	10885	10879
Number of clusters	264	264	264	264	264
Mean dep. var. control group	184.052	172.693	190.234	186.477	185.329
SD dep. var. control group	41.081	46.528	52.637	49.517	42.864
5th grade – Primary schools					
Treatment	-0.068	-0.067	-0.091	-0.070	-0.074
	(0.087)	(0.097)	(0.091)	(0.087)	(0.084)
	[0.584]	[0.641]	[0.407]	[0.531]	[0.417]
Number of observations	3179	2885	2885	2977	2978
Number of clusters	92	92	92	92	92
Mean dep. var. control group	157.452	157.540	173.368	154.288	149.499
SD dep. var. control group	36.022	43.798	60.456	37.359	28.700
6th grade – Lower secondary schools					
Treatment	0.146**	0.177**	0.158**	0.103*	0.123**
	(0.061)	(0.073)	(0.075)	(0.054)	(0.062)
	[0.034]	[0.026]	[0.057]	[0.070]	[0.054]
Number of observations	4511	4014	4013	4134	4131
Number of clusters	99	99	99	99	99
Mean dep. var. control group	162.845	151.930	172.451	160.075	170.685
SD dep. var. control group	31.523	42.024	47.502	35.775	35.164
10th grade – Upper secondary schools					
Treatment	-0.011	-0.015	-0.016	-0.026	0.051
	(0.078)	(0.088)	(0.112)	(0.062)	(0.053)
	[0.884]	[0.847]	[0.878]	[0.610]	[0.417]
Number of observations	5070	4467	4467	3774	3770
Number of clusters	73	73	73	73	73
Mean dep. var. control group	215.446	198.009	214.086	233.701	223.680
SD dep. var. control group	26.923	38.838	41.371	26.369	23.650

Note: *Significant at 10 percent; **Significant at 5 percent; ***Significant at 1 percent using conventional inference (i.e., not adjusting for multiple hypothesis correction). Unit of observation: student. Outcome variables in the column headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level are in parentheses. Romano-Wolf step-down adjusted p-values robust to multiple hypothesis testing across grades are estimated using 10,000 bootstrap replications with stratified and clustered resampling and reported in brackets. The coefficients are expressed in terms of standard deviations from the control group, while the unconditional mean and standard deviation of the dependent variable refer to the raw values in the control group.

Table 2: Impact on Student Progression Rates

	Grade level				*Student level*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
					Passing			
Treatment	4.70**	2.44	8.46**	2.46	4.51**	1.29	7.00**	4.25
	(1.83)	(2.55)	(3.30)	(3.61)	(2.21)	(2.65)	(3.10)	(4.13)
		[0.548]	[0.032]	[0.548]		[0.520]	[0.016]	[0.335]
Number of observations	277	95	104	78	17276	3629	5490	8157
Number of clusters					277	95	104	78
Mean dep. var. control group	70.97	83.55	63.56	66.22	59.91	79.60	58.73	52.81
SD dep. var. control group	18.04	13.64	17.05	16.23	49.01	40.31	49.24	49.93
					Dropout			
Treatment	-0.20	-0.16	-1.61	1.60	-0.85	0.26	-4.35**	1.13
	(0.83)	(0.79)	(1.27)	(2.21)	(1.39)	(1.38)	(1.82)	(2.67)
		[0.845]	[0.492]	[0.714]		[0.822]	[0.009]	[0.822]
Number of observations	277	95	104	78	17290	3637	5494	8159
Number of clusters					277	95	104	78
Mean dep. var. control group	6.19	2.09	6.84	10.17	16.83	8.19	13.55	22.40
SD dep. var. control group	7.96	3.87	7.15	10.17	37.42	27.43	34.23	41.70
					Repetition			
Treatment	-4.49***	-2.28	-6.85**	-4.06	-3.66**	-1.55	-2.65	-5.38*
	(1.70)	(2.38)	(2.91)	(3.61)	(1.69)	(1.87)	(2.81)	(2.97)
		[0.445]	[0.053]	[0.445]		[0.390]	[0.390]	[0.066]
Number of observations	277	95	104	78	17276	3629	5490	8157
Number of clusters					277	95	104	78
Mean dep. var. control group	22.84	14.37	29.59	23.61	23.25	12.20	27.72	24.78
SD dep. var. control group	15.27	12.86	14.91	13.71	42.25	32.74	44.77	43.18

Note: *Significant at 10 percent; **Significant at 5 percent; ***Significant at 1 percent using conventional inference (i.e., not adjusting for multiple hypothesis correction). School-level data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses, respectively. Unit of observation: school and student. Outcome variables in the panel headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors for school-level regressions and standard errors clustered at the school level for student-level regressions in parentheses. Romano-Wolf stepdown adjusted p-values robust to multiple hypothesis testing across grades are estimated using 10,000 bootstrap replications with stratified (and clustered resampling for student-level regressions) and reported in brackets. The coefficients are expressed in terms of percentage points. The mean and standard deviation of the dependent variable in the control group are unconditional.

5 Potential Mechanisms and Heterogeneous Results

The results show that PIP had meaningful impacts on student learning and progression, but results are concentrated in 6th grade. In this section, we explore potential mechanisms through which the program may have affected these outcomes and assess whether they can explain the treatment effect heterogeneity by grade.

The program's key components— teacher autonomy, pedagogical innovation, technical support, and financial resources—were designed to affect student outcomes through two main, potentially complementary, mechanisms:

1. Increasing teacher motivation and engagement through the provision of autonomy and resources to develop their own project with technical support;
2. Enhancing students' motivation and strengthening student-teacher interactions through the implementation of innovative pedagogical projects in schools.

We hypothesize that teacher autonomy over the design of the projects and use of the grant can affect outcomes through both mechanisms, as it may crowd in teacher intrinsic motivation and lead to better locally tailored projects. The second mechanism suggests that innovative projects, aimed at strengthening student-teacher interactions, could generate positive results regardless of teacher autonomy.

Although we did not collect information on some of the hypothesized impact mediators, such as teacher autonomy and motivation or the quality of teacher-student interactions, we present suggestive evidence using non-experimental variation in each component, qualitative evidence from the implementers, and the existing literature.

5.1 Impact on Teacher Engagement

We proxy teachers' engagement by their decision to remain in the school, measured as teacher retention (see Section 3.2).[36] Low teacher retention might affect students'

[36] We note that there are other channels, besides enhanced autonomy and motivation, that we do not directly observe through which the program might affect teacher retention. For example, if the program

achievement and motivation, particularly in 6th grade. Mentors stated that 6th graders face significant shocks when transitioning between levels of education. In primary school, students have a single teacher, which allows for a close student-teacher relationship. These ties are weaker in lower secondary education, as students have multiple teachers (at least five in our sample). The potential negative impact of this transition is well documented in the United States (Bedard and Do, 2005; Cook et al., 2008; Hanewald, 2013) and has been recently investigated in Brazil (Santos et al., 2017). The latter study evaluates the impact of a pilot in municipal schools in Rio de Janeiro, Brazil, which expanded primary schools to include grade 6. They find that having 6th grade in the primary school increases learning by 0.16 SD and they provide suggestive evidence that strengthened student-teacher relationships mediated some of the effects on learning. This might imply that enhancing teacher retention in 6th grade could be particularly effective in improving student outcomes, in line with recent evidence from Ronfeldt et al. (2013), Fagernäs and Pelkonen (2020), Zeitlin (2021), and Akhtari et al. (2022)—the latter in the Brazilian context.

Pooling all grades, the ITT estimate on teacher retention is positive but not statistically significant (column 1 in Panel 1 of Table 3). Splitting the analysis by grade, we find that teacher retention increased by 6.4 pp in grade 6, i.e., a 20.7 percent increase in teacher retention over the control mean of 30.9 percent (column 3).[37,38] We also recover TOT estimates for participating teachers as the program did not require all teachers to participate and the decision for which and how many teachers to include was at the discretion of the schools. Since primary schools have class-specific teachers, while secondary schools have subject-specific teachers, teacher participation was mechanically higher in 5th grade. We find similar rates of participation in 6th and 10th grades, where roughly one third of the teachers complied with treatment assignment (Panel 1 of online

was perceived as a signal that the government has increased the degree of resources that will be provided to the school. We explore the role of the grant in the mechanism section.

[37]The coefficient remains statistically significant at conventional levels after adjusting for multiple hypothesis testing and when using alternative average treatment effect estimators (online Appendix Table E10).

[38]The lower teacher retention in control schools is driven by more teachers moving to other schools rather than leaving the state education system.

Appendix Table D6).[39] In 6th grade, the TOT estimate for teacher retention is equal to 19 pp, or 61 percent over the control mean (Panel 2 of online Appendix Table D6). Moreover, we find that observable differences between participant and non-participant teachers are small in magnitude and mostly not distinguishable from zero (online Appendix Table B2). This suggests that the differential selection across grades along these characteristics is not likely driving the heterogeneous results.

Since 6th grade has the largest teacher turnover at baseline, we explore first whether the impact on teacher retention varies with baseline levels of turnover. We define 'Low teacher retention' at the grade level as a dummy that equals one if the school has a retention rate below the sample median of that grade in the year before the intervention, and zero otherwise. Panel B in Table 3 suggests that the increase in teacher retention is concentrated in schools with low teacher retention rates at baseline (column 1): in these schools, the program increased teacher retention by 10.1 pp, i.e., a 15.5 percent increase in teacher retention over the control mean of 35 percent, almost closing the gap in teacher retention between low and high-retention schools, across grades. Generally, impacts on retention seem limited to low-retention schools, and the same patterns and magnitudes are observed in both 6th and 10th grades (columns 3 and 4, respectively).

In the next sub-section, we combine the results on teacher retention and implementation to verify whether we observe similar patterns on student outcomes.

5.2 Implementation of Innovative Pedagogical Projects

Project overall quality. We first explore whether the heterogenous results by grade are driven by better-designed projects in grade 6. However, we find no differences across grades in the average quality of proposals, measured by their score (online Appendix Figure B7). Moreover, these scores do not seem to have any predictive power on either the rate of project implementation or student outcomes, such as progression and learning

[39] In our sample, there are an average of 1.9 teachers in 5th grade, 7.5 in 6th grade, and 11.6 in 10th grade. This translates to 1.3, 3, and 5.6 teachers participating in PIP, on average, in grades 5, 6, and 10, respectively.

Table 3: Impact on Teacher Retention

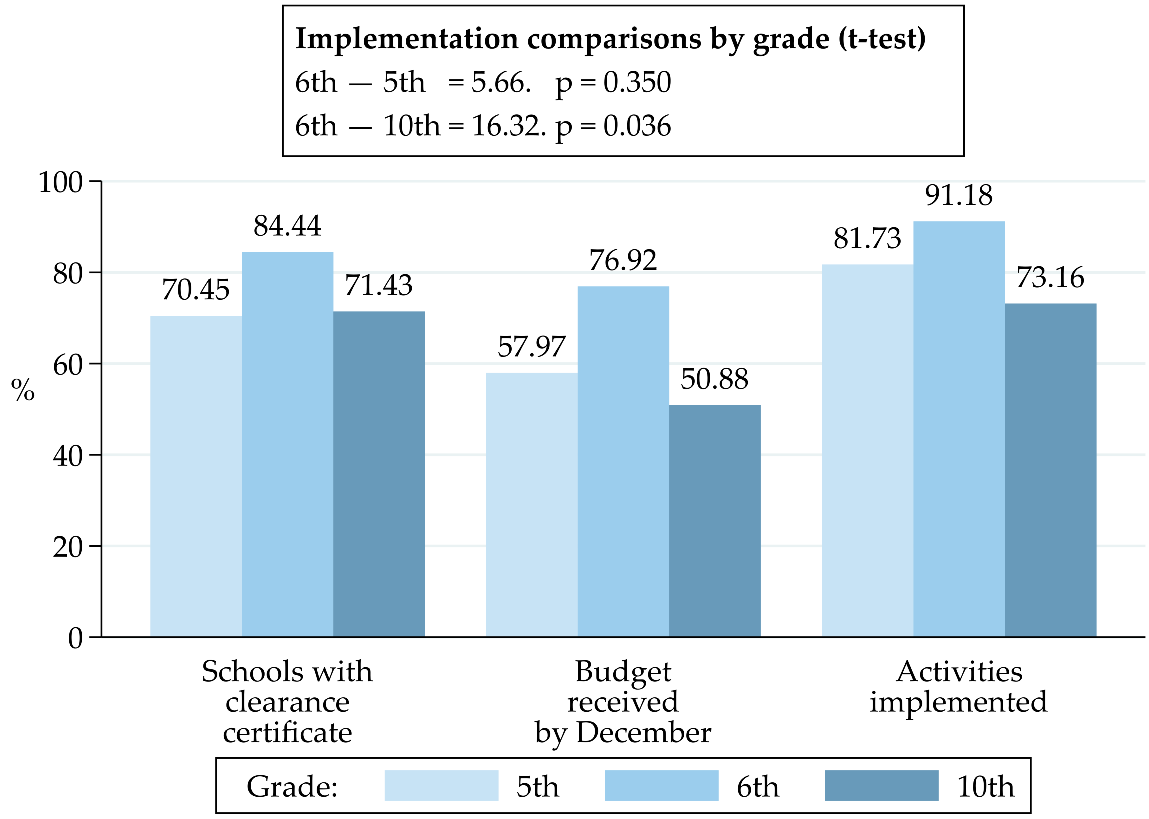
	(1) All	(2) 5th	(3) 6th	(4) 10th
Panel A – Overall impact				
Treatment	0.036	-0.064	0.064*	0.033
	(0.029)	(0.065)	(0.037)	(0.049)
		[0.368]	[0.069]	[0.414]
Number of observations	1882	189	784	909
Number of clusters	277	95	104	78
Mean dep. var. control group	0.709	0.761	0.691	0.714
SD dep. var. control group	0.454	0.428	0.463	0.452
Panel B – Impact by retention at baseline				
Treatment	-0.030	-0.187*	0.014	-0.038
	(0.040)	(0.110)	(0.046)	(0.064)
Treatment × Low teacher retention rate at baseline	0.130**	0.191	0.104	0.139
	(0.057)	(0.137)	(0.074)	(0.092)
Low teacher retention rate at baseline	-0.118***	-0.069	-0.127**	-0.115*
	(0.036)	(0.083)	(0.050)	(0.058)
Constant	0.776***	0.851***	0.756***	0.780***
	(0.023)	(0.067)	(0.032)	(0.036)
Total effect on schools with low retention at baseline: Treatment + Treatment × low-retention dummy				
$\sum \hat{\beta}$	0.101	0.004	0.118	0.101
P-value	0.016	0.960	0.043	0.131
Unconditional mean of the dependent variable in the control group:				
Schools with low retention at baseline	0.650	0.786	0.626	0.649
Schools with high retention at baseline	0.780	0.833	0.762	0.786

Note: *Significant at 10 percent; **Significant at 5 percent; ***Significant at 1 percent using conventional inference (i.e., not adjusting for multiple hypothesis correction). Data are from Rio Grande do Norte 2016 and 2017 teacher censuses. Unit of observation: teacher. $\sum \hat{\beta}$ is the sum of the treatment effect with the interaction variable coefficient. The p-value refers to the null hypothesis $\sum \hat{\beta} = 0$. All regressions are linear probability model with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level in parentheses. Romano-Wolf stepdown adjusted p-values robust to multiple hypothesis testing across grades are estimated using 10,000 bootstrap replications with stratified (and clustered resampling for student-level regressions) and reported in brackets. Note that the median teacher retention at baseline for 5th grade is equal to 1, so 'Low teacher retention rate at baseline' indicates any level of retention below 1. This is due to the fact that, in most schools, 5th grade has only one teacher, thus the school retention rate variable is either equal to 0 or 1.

(online Appendix Table D7 and D8, respectively).

Rate of overall project implementation in schools. We compare administrative data to assess whether project implementation varied across grades. We report three measures of school-level implementation: (i) obtaining the clearance certificate, which is a necessary requirement for schools to receive funding from any state-level educational program;[40] (ii) percentage of project funds received by the end of the school year; and (iii) percentage of planned activities that were implemented by the end of the year. We observe substantial differences in rates of implementation across the three grades (Figure 1). Each of the indicators shows higher rates in 6th grade, on average 83 percent of activities were implemented. Moreover, we find that implementation was particularly poor in 10th-grade schools with low teacher retention, where the average share of activities implemented was only 60 percent (online Appendix Figure B8). These results are in line with the positive and larger impacts we find in grade 6.

Figure 1: Implementation by Grade



Note: The bars show the unconditional mean of the three implementation measures defined in Subsection 5.5. Comparisons by grade are done on the third measure using region fixed effects and robust standard errors. *P*-values based on standard *t*-test. 'Planned activities implemented' are defined as the ratio of the number of activities that were implemented over the number of planned activities described in the work plan. Data are from the State Secretariat of Education (SEE). Sample: schools in the treatment group.

[40]We indeed find that obtaining the clearance certificate is what most predicts the rate of implementation (online Appendix Table D7). We find that being assigned to receive the treatment increases the likelihood of schools obtaining the clearance certificate during the year of the intervention by 41 pp. This impact does not differ by grade (online Appendix Table D9). By the end of the year, all schools had the clearance certificate and were therefore entitled to receive the grant transfer.

5.3 Role of Teacher Retention and Project Implementation

We first investigate whether the increase in teacher retention described in the previous section is driving the results, by exploiting the fact that many 6th-grade teachers also teach other grades, where no innovative pedagogical projects are implemented. According to the school census data, 90.4 percent of 6th-grade teachers also teach in 7th grade, 81.8 percent teach 8th grade, and 73.2 percent teach 9th grade.[41] As a result, the reduction in 6th-grade turnover also mechanically affects turnover in the other grades in the same schools (Panel A in online Appendix Table D10). We compare student-level outcomes for 6th grade schools in their remaining lower-secondary grades (Panel B in online Appendix Table D10).[42] We only have access to data on student progression in other grades, as the standardized test was not implemented in upper grades. We find no indication of positive spillovers to other grades, which might suggest increasing teacher retention alone might not be sufficient to affect student outcomes: positive results in 6th grade are likely driven by the combination of an increase in teacher retention and the implementation of the innovative projects.[43] On the other hand, we find no negative spillovers on other grades, which suggests that teachers did not increase effort in 6th grade at the cost of other grades.

Further, we explore whether the observed patterns of impacts on teacher retention, being concentrated in schools with low teacher retention at baseline, and implementation, are also observed for child outcomes. In grade 6, we observe no difference in in-school implementation by baseline levels of teacher retention (online Appendix Figure B8), while for grade 10 implementation is substantially worse in schools with low teacher retention at baseline. Impacts on progression do not differ by baseline levels of teacher retention, while learning gains are concentrated among schools with low retention at baseline, where teachers are most affected (first panel of online Appendix Table D13). This is in line with

[41] The percentage is balanced between treatment and control schools.

[42] The results using grade-level data from SIGEduc are very similar and are presented in online Appendix Table D11.

[43] Results by teacher retention at baseline also show no impacts on other grades (online Appendix Table D12).

strong, positive correlations between student learning outcomes and teacher retention in the control group, while correlations with student progression are more mixed. This might suggest that teacher retention, and engagement, may be critical to generating learning gains through the projects. In grade 10 schools with low teacher retention at baseline, teachers were impacted but implementation of projects was particularly poor, possibly explaining the lack of impacts on learning (second panel of online Appendix Table D13).

5.4 Student-Teacher Interactions

Changes in teacher-student interactions. We explore descriptively whether learning results are driven by the educational content of the projects or by changes in pedagogy and student-teacher interactions, which can also be induced by increased teacher engagement. While 70 percent of projects implemented in 6th grade were focused on reading and literacy (online Appendix Figure A1b), we find ITT effects of similar magnitude in all subjects tested. Moreover, the standardized assessment was not high-stakes for students, hence we do not expect teachers to have content tailored to the exam. At the same time, the correlation between scores on different subjects is not particularly high, ranging from 0.34 (human and natural sciences) to 0.57 (math and Portuguese). This limits the likelihood that the impact is explained by the curricular content of the activities only.

Socio-Emotional Skills. Throughout the development of the projects, teachers were encouraged to design an intervention that would change student-teacher interactions and engage students by exposing them to learning opportunities outside the classroom, moving away from traditional, lecture-based teaching. As a consequence, resulting projects may have affected students' socio-emotional skills directly and indirectly. Directly, by strengthening teacher-student interactions in innovative pedagogical activities. Indirectly, through the complementarities between cognitive and non-cognitive skills.

Pooling all grades, we find that the program had a positive effect on conscientiousness and extroversion (Table 4). Among the Big Five, the trait of "conscientiousness" is commonly associated with the acquisition of cognitive skills (Poropat, 2009; Ivcevic and Brackett, 2014). It encompasses traits such as self-control, organization, responsibility, and perseverance. The point estimates are statistically significant at conventional levels. However, in line with previous results, these appear to be driven by the impacts on 6th graders (0.17 SD and 0.21 SD, respectively), which both pass multiple hypothesis testing.[44]

We observe that student test scores and socio-emotional skills are positively correlated in the tested sample at endline, which points to possible complementarities between socio-emotional and cognitive development,[45] regardless of treatment status (online Appendix Figure D4). Unfortunately, as mentioned in Section 3.2, data on socio-emotional skills were only collected for a random subset of students in each school. When we restrict the sample to the subset of students who took the socio-emotional test, we are unable to detect significant effects of the program on learning outcomes, therefore we cannot further investigate the mediating role of socio-emotional skills on learning outcomes or vice versa.

5.5 Role of Complementary Program Components

Complementing management capacity. Financial resources do not seem to be the main binding constraint for the procurement of school supplies and implementation of new pedagogical activities in our context. Data on disbursement rates of public funds, such as *Programa Dinheiro Direto na Escola* (PDDE) run by *Fundo Nacional de Desenvolvimento da Educao* (FNDE) reveal that, on average, only 30.5 percent of federal funds assigned to schools to invest in infrastructure and pedagogy are spent.[46] Rather than solving budgetary scarcity, the program might have prompted schools to overcome administrative

[44]Correcting for non-random sample selection at the school level mostly does not invalidate our results (online Appendix Table C7). Treatment effect bounds are computed at the school level given that differential participation by treatment assignment is mostly driven by whole schools not having taken the test. The other robustness checks are in online Appendix E.

[45]These complementarities might not be universal (Laajaj and Macours, 2019).

[46]The amount of operating funds the school can receive is based on the number of students measured by the school census of the previous year. The school receives it in two installments (April and September). A key issue is that the school must deliver the clearance certificate from the previous year to be eligible to receive the funds.

Table 4: Impact on Socio-Emotional Skills

	(1) Agreeableness	(2) Conscientiousness	(3) Extroversion	(4) Neuroticism	(5) Openness
All schools					
Treatment	0.048	0.115**	0.116**	0.037	0.058
	(0.056)	(0.054)	(0.054)	(0.047)	(0.054)
Number of observations	3560	3560	3560	3558	3560
Number of clusters	235	235	235	235	235
Mean dep. var. control group	4.413	4.331	4.199	4.007	4.105
SD dep. var. control group	0.975	1.053	0.777	0.738	0.970
5th grade – Primary schools					
Treatment	0.023	0.094	0.049	-0.019	-0.061
	(0.097)	(0.095)	(0.097)	(0.073)	(0.094)
	[0.798]	[0.365]	[0.511]	[0.743]	[0.400]
Number of observations	1296	1296	1296	1294	1296
Number of clusters	85	85	85	85	85
Mean dep. var. control group	4.468	4.359	4.287	4.040	4.193
SD dep. var. control group	1.049	1.108	0.851	0.738	0.997
6th grade – Lower secondary schools					
Treatment	0.078	0.173*	0.208**	0.058	0.139
	(0.099)	(0.098)	(0.097)	(0.092)	(0.097)
	[0.689]	[0.073]	[0.021]	[0.665]	[0.198]
Number of observations	1270	1270	1270	1270	1270
Number of clusters	87	87	87	87	87
Mean dep. var. control group	4.390	4.265	4.156	3.971	3.950
SD dep. var. control group	1.090	1.176	0.858	0.770	1.089
10th grade – Upper secondary schools					
Treatment	0.042	0.069	0.085	0.082	0.110
	(0.091)	(0.080)	(0.079)	(0.075)	(0.080)
	[0.798]	[0.365]	[0.311]	[0.409]	[0.198]
Number of observations	994	994	994	994	994
Number of clusters	63	63	63	63	63
Mean dep. var. control group	4.378	4.387	4.152	4.017	4.212
SD dep. var. control group	0.663	0.761	0.514	0.692	0.701

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. Outcome variables in the column headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level in parentheses. Romano-Wolf stepdown adjusted p-values robust to multiple hypothesis testing across grades are estimated using 10,000 bootstrap replications with stratified and clustered resampling and reported in brackets. The coefficients are expressed in terms of standard deviations from the control group, while mean and standard deviation of the dependent variable refer to the raw values in the control group. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score.

bottlenecks in access to school funding, for instance by obtaining the clearance certificate. This might have led to increased spending across the board, which in turn supported principals to spend available public funds more effectively. Online Appendix Table D15 shows that this was the case for schools participating in PIP with their 6th grade: treated schools disbursed a 60 pp higher share of PDDE funds compared to the control in the year following the intervention. We do not observe this in the school year we evaluate, but this does suggest that the provision of technical support in project management and procurement is important and can make a substantial difference in the implementation of planned activities.

Quality of technical assistance. We note that it is unlikely that, by chance, better mentors were assigned to support 6th grade: all mentors worked across grades (online Appendix Figure B9). We observe that while mentors were more likely to visit schools participating in PIP with their 6th grade, the difference is minimal, with 0.25 extra visits compared to 5th grade (online Appendix Figure B10). During a focus group discussion with mentors, we found that they had more experience with teaching and implementing projects in lower grades, which may have resulted in the technical assistance being better tailored to grades 5 and 6.

Role of the grant. Differences in grant size are not likely driving the heterogeneity in implementation. Since grant size is determined by the number of classes instead of by the number of students, the value per student may differ across grades. Due to typically smaller classes,[47] grade 5 and grade 6 schools received BRL 456 more per student, or USD 114 than grade 10 schools did (online Appendix Figure B11), which is almost double. Despite these differences, we find that the grant amount per student does not have any predictive power on project implementation or on students' outcomes in all grades (online Appendix Table D7 and D14). These should not be interpreted as causal impacts of the grant as the allocated value within the treatment group was not random.

5.6 Heterogeneity in Other Grade Characteristics

We explore whether heterogeneity in baseline grade characteristics might explain the differential impact in 6^{th} grade. We start by noticing that the state administers the

[47]This difference is common to the study context and is not an artifact of the program selection of eligible state schools nor the grade targeting rule.

universe of schools offering upper-secondary education and only a smaller share of schools offering lower-secondary education. We compare observed characteristics across grades to assess whether the 6^{th} grade state schools are different from schools offering either the 5^{th} or 10^{th} grades in the study sample.

Schools that were targeted in their 6th grade are, on average, larger than schools targeted in 5th grade, but smaller than schools targeted in 10th grade, in terms of number of personnel, teachers, students, and classes (online Appendix Table B3). In addition, student-teacher ratios are higher in grade 10, namely 28.1 versus 23.8 in grade 6. This is the case both at the school and the grade level. Importantly for external validity, the observed pattern in terms of school size represents a general characteristic of this grade and it is not a specific feature of our sample of state schools. On the other hand, teacher background and turnover are similar across grades and administration types.

6 Policy Analysis

In this section, we use the main results on learning and progression to produce back-of-the-envelope estimates for the impact of the program on school quality indicators and individuals' expected earnings.

Quality of Education. We use the ITT estimates to compute the counterfactual distribution of two national quality of education indicators. First, online Appendix Figure F1 shows that if students retain their learning gains over time, as measured by SAEB scores, the impact of the PIP would close half of the knowledge gap between RN and the country's average by the end of grade 9. Second, combining impacts on progression and learning suggests that the PIP would help RN state schools move upwards in the IDEB ranking by at least two positions (online Appendix Figure F2).[48] This strategy is described in more detail in the online Appendix F.1.

[48]This analysis assumes no effect on teacher selection. However, it might be that more qualified teachers participate in the program and, because of it, end up remaining in their school and foregoing the opportunity to move to a better, or just more suitable, institution. In online Appendix Table F2, we test for differential retention and do not detect any heterogeneous treatment effect by qualification.

Expected Returns to Education. We expect the intervention to impact labor market outcomes of the 6th graders in the long term through two channels: first via learning gains among those who stayed in school (productivity channel), and second, via the higher probability of remaining in school conditional on passing grade 6 (a combination of productivity effects with signaling or diploma effects). The first channel focuses on the improved quality of education, while the second reflects extra years of education among more knowledgeable students.

Using the ITT effects of the PIP on learning as being approximately equal to 0.5 extra years of schooling, a back-of-the-envelope calculation suggests an NPV on future earnings of BRL 29,148.97 or USD 7,287.24. The second channel is through the increase in student years of schooling through a reduction in repetition which we estimate leads to about 0.4 extra years of schooling, with an NPV on future earnings of BRL 23,319.18 (or USD 5,829.79). The full effect on expected earnings would then range from USD 7,000 to 13,000, or 28 to 52 times the annual Brazilian minimum wage. The data and methodology used for the calculations are described in online Appendix F.2.

This calculation assumes all the expected impacts on future earnings are driven by direct or indirect gains in learning. However, besides mediating the accumulation of cognitive skills, there may be direct impacts of socio-emotional skills on labor market outcomes, making this a lower-bound estimate.

7 Conclusion

In this paper, we studied the impacts of an education program implemented in state schools of Rio Grande do Norte, one of the poorest Brazilian states. The intervention combined teacher autonomy, technical assistance, and targeted school funds to increase students' motivation and learning and strengthen teacher-student engagements. In doing so, the program leveraged teachers' local knowledge to design and implement a tailored set of innovative pedagogical activities.

We found a positive impact of 6.6 percent on progression rates, with the effects being concentrated in grade 6 (12 to 13 percent increase over the control mean). We detected a meaningful positive impact on learning outcomes for 6th graders only. The positive effects on math and reading scores point to an impact equivalent to almost half a year of additional schooling.

We found suggestive evidence that the program increased teachers' and students' motivation and engagement, but mainly in grade 6. This result is consistent with other findings showing that strengthening student-teacher interactions during the transition from elementary to lower-secondary can impact students' dropout rates, and learning. Data on the program's implementation suggest that part of the larger effects on grade 6 might be related to a better implementation of the program. Anecdotal evidence collected from focus group discussions with both the school mentors and the members of the state Secretariat of Education suggested that part of the program's success in grade 6 is related to the fact that lower-secondary education does not get as much attention and resources as the early and later K-12 grades.

Program components were not randomized so we cannot tell apart the relative importance of each component and the exact channel through which the program impacted teacher and student outcomes. Second, data constraints limit the extent of analysis along the causal impact chain. Despite this, we provide suggestive evidence using non-experimental variation in each component, anecdotal reports from implementers, and the existing literature. The lack of results in other grades may be explained by lower rates of implementation or the approach being particularly appropriate in a context where the motivation of agents and final recipients, in this case, students, is essential to affect outcomes. It appears that none of PIP's components alone can explain the results, in line with the still scant evidence that school policies designed to tackle multiple constraints at once are more likely to work. More research is needed to understand in which settings this decentralized approach is more likely to succeed and to further explore the potential complementarities between teacher autonomy, technical assistance, and project-based funding.

Overall, our results show that significant improvements to educational outcomes can be obtained by leveraging local staff knowledge and other existing school resources. These findings have direct implications for policy design in countries that might provide public basic education without having either fiscal space to design pay-for-performance schemes at scale and/or effective monitoring mechanisms in place.

References

Akhtari, M., D. Moreira, and L. Trucco (2022): "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services," *American Economic Review*, 112, 442–493.

Almeida, R., A. Bresolin, B. Borges, K. Mendes, and N. Menezes-Filho (2016): "Assessing the Impacts of Mais Educação on Educational Outcomes: Evidence between 2007 and 2011," Policy Research Working Paper No. 7644, The World Bank: Washington, DC.

Alves, M. T. G., J. F. Soares, and F. P. Xavier (2016): "Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais," *Revista Brasileira de Sociologia*, 4, 49–82.

Andrabi, T., N. Bau, J. Das, N. Karachiwalla, and A. Khwaja (2021): "Crowding in Private Quality: The Equilibrium Effects of Public Spending in Education," Working Paper.

Angrist, N., S. Djankov, P. K. Goldberg, and H. A. Patrinos (2021): "Measuring Human Capital Using Global Learning Data," *Nature*, 592, 403–408.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016): "Teacher Quality and Learning Outcomes in Kindergarten," *The Quarterly Journal of Economics*, 131, 1415–1453.

Bandiera, O., M. C. Best, A. Q. Khan, and A. Prat (2021): "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," *The Quarterly Journal of Economics*, 136, 2195–2242.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017): "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application," *Journal of Economic Perspectives*, 31, 73–102.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton (2016): "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India," NBER Working Paper No. 22746.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2021): "Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training," *American Economic Journal: Economic Policy*, 13, 36–66.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007): "Remedying Education: Evidence from Two Randomized Experiments in India," *The Quarterly Journal of Economics*, 122, 1235–1264.

Bau, N. and J. Das (2020): "Teacher Value Added in a Low-Income Country," *American Economic Journal: Economic Policy*, 12, 62–96.

Beasley, E. and E. Huillery (2017): "Willing but Unable? Short-Term Experimental Evidence on Parent Empowerment and School Quality," *The World Bank Economic Review*, 31, 531–552.

Bedard, K. and C. Do (2005): "Are Middle Schools More Effective? The Impact of School Structure on Student Outcomes," *Journal of Human Resources*, 40, 660–682.

Beg, S., W. Halim, A. M. Lucas, and U. Saif (2022): "Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not," *American Economic Journal: Economic Policy*, 14, 61–90.

Blimpo, M. P., M. Blimpo, D. Evans, and N. Lahire (2015): "Parental Human Capital and Effective School Management: Evidence from the Gambia," Policy Research Working Paper No. 7238, The World Bank: Washington, DC.

Burgess, R., M. Hansen, B. A. Olken, P. Potapov, and S. Sieber (2012): "The Political Economy of Deforestation in the Tropics," *The Quarterly Journal of Economics*, 127, 1707–1754.

Carneiro, P., O. Koussihouèdé, N. Lahire, C. Meghir, and C. Mommaerts (2020): "School Grants and Education Quality: Experimental Evidence from Senegal," *Economica*, 87, 28–51.

Carvalho, J. S. d. O. (2016): "O Projeto de Inovao Pedaggica (PIP) e as Prticas Inovadoras dos Professores da Rede Estadual do Ensino Mdio do RN," Universidade Estadual do Rio Grande do Norte, Master's thesis.

Cassar, L. and S. Meier (2018): "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, 32, 215–38.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104, 2633–79.

Cook, P. J., R. MacCoun, C. Muschkin, and J. Vigdor (2008): "The Negative Impacts of Starting Middle School in Sixth Grade," *Journal of Policy Analysis and Management*, 27, 104–121.

Cunha, F. and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman (2013): "School Inputs, Household Substitution, and Test Scores," *American Economic Journal: Applied Economics*, 5, 29–57.

Deci, E. L. (1971): "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology*, 18, 105.

Deci, E. L. and R. M. Ryan (1985): *Intrinsic Motivation and Self Determination in Human Behaviour*, Plenum, New York.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2018): "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India," *Econometrica*, 86, 2123–2160.

Evans, D. K. and F. Yuan (2019): "Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms," Policy Research Working Paper No. 8752, The World Bank: Washington, DC.

Fagernäs, S. and P. Pelkonen (2020): "Teachers, Electoral Cycles, and Learning in India," *Journal of Human Resources*, 55, 699–732.

Firpo, S., M. N. Foguel, and H. Jales (2020): "Balancing Tests in Stratified Randomized Controlled Trials: A Cautionary Note," *Economics Letters*, 186, 108771.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973.

Gibbons, C. E., J. C. S. Serrato, and M. B. Urbancic (2018): "Broken or Fixed Effects?" *Journal of Econometric Methods*, 8.

Gilligan, D. O., N. Karachiwalla, I. Kasirye, A. M. Lucas, and D. Neal (2022): "Educator Incentives and Educational Triage in Rural Primary Schools," *Journal of Human Resources*, 57, 79–111.

Glewwe, P., M. Kremer, and S. Moulin (2009): "Many Children Left Behind? Textbooks and Test Scores in Kenya," *American Economic Journal: Applied Economics*, 1, 112–35.

Glewwe, P. and K. Muralidharan (2016): "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications," in *Handbook of the Economics of Education*, Elsevier, vol. 5, 653–743.

Gray-Lobe, G., A. Keats, M. Kremer, I. Mbiti, and O. W. Ozier (2022): "Can Education Be Standardized? Evidence from Kenya," *Evidence from Kenya (June 5, 2022). University of Chicago, Becker Friedman Institute for Economics Working Paper.*

Hanewald, R. (2013): "Transition between Primary and Secondary School: Why It Is Important and How It Can Be Supported." *Australian Journal of Teacher Education*, 38, n1.

Hanushek, E. A. and L. Woessmann (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–68.

Imbens, G. W. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Instituto Brasileiro de Geografia e Estatística (IGBE) (2018): "Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua," Rio de Janeiro.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (2015): "Censo Escolar," Ministério da Educação: Brasília.

Ivcevic, Z. and M. Brackett (2014): "Predicting School Success: Comparing Conscientiousness, Grit, and Emotion Regulation Ability," *Journal of Research in Personality*, 52, 29–36.

Jackson, C. K. (2018): "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes," *Journal of Political Economy*, 126, 2072–2107.

Kautz, T., J. J. Heckman, R. Diris, B. Ter Weel, and L. Borghans (2014): "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success," NBER Working Paper No. 20749.

Laajaj, R. and K. Macours (2019): "Measuring Skills in Developing Countries," *Journal of Human Resources*, 1018–9805R1.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 76, 1071–1102.

Marinelli, H. A., S. Berlinski, and M. Busso (2021): "Remedial Education: Evidence from a Sequence of Experiments in Colombia," *Journal of Human Resources.*

Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2019): "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania," *The Quarterly Journal of Economics*, 134, 16271673.

McEwan, P. J. (2015): "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments," *Review of Educational Research*, 85, 353–394.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019): "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India," *American Economic Review*, 109, 1426–1460.

Oliveira, L. F. B., R. Terra, et al. (2016): "Impacto do Programa Mais Educação em Indicadores Educacionais," IPC-IG Working Paper n. 147, International Policy Centre for Inclusive Growth: Brasília.

Organisation for Economic Co-operation and Development (OECD) (2015): "PISA: Programme for International Student Assessment," OECD Education Statistics, OECD Publishing: Paris.

——— (2016): *Education at a Glance 2016*, OECD Publishing: Paris.

Poropat, A. E. (2009): "A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance," *Psychological Bulletin*, 135, 322.

PRIMI, R., C. ZANOS, D. SANTOS, F. DE FRUYT, AND O. P. JOHN (2016): "Anchoring Vignettes Can They Make Adolescent Self-Reports of Social-Emotional Skills More Reliable, Discriminant, and Criterion-Valid?" *European Journal of Psychological Assessment*, 32.

RASUL, I. AND D. ROGGER (2018): "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *The Economic Journal*, 128, 413–446.

RASUL, I., D. ROGGER, AND M. J. WILLIAMS (2021): "Management, Organizational Performance, and Task Clarity: Evidence from Ghanas Civil Service," *Journal of Public Administration Research and Theory*, 31, 259–277.

ROGGER, D. AND R. SOMANI (2018): "Hierarchy and Information," Policy Research Working Paper No. 8595, The World Bank: Washington, DC.

ROMANO, J. P. AND M. WOLF (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73, 1237–1282.

——— (2016): "Efficient Computation of Adjusted p-Values for Resampling-Based Stepdown Multiple Testing," *Statistics & Probability Letters*, 113, 38–40.

RONFELDT, M., S. LOEB, AND J. WYCKOFF (2013): "How Teacher Turnover Harms Student Achievement," *American Educational Research Journal*, 50, 4–36.

SANTOS, D. D., L. G. SORZAFAVE, A. C. NICOLELLA, AND E. G. SANT'ANNA (2017): "Mais é menos? O impacto do Projeto 6º Ano Experimental–SME/RJ," *Estudos em Avaliação Educacional*, 28, 718–747.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO) (2020): "UNESCO Institute for Statistics, National Monitoring," Paris.

WORLD BANK (2018): *World Development Report 2018: Learning to Realize Education's Promise*, The World Bank: Washington, DC.

ZEITLIN, A. (2021): "Teacher Turnover in Rwanda," *Journal of African Economies*, 30, 81–102.

ZUCKERMAN, M., J. PORAC, D. LATHIN, AND E. L. DECI (1978): "On the Importance of Self-Determination for Intrinsically-Motivated Behavior," *Personality and Social Psychology Bulletin*, 4, 443–446.

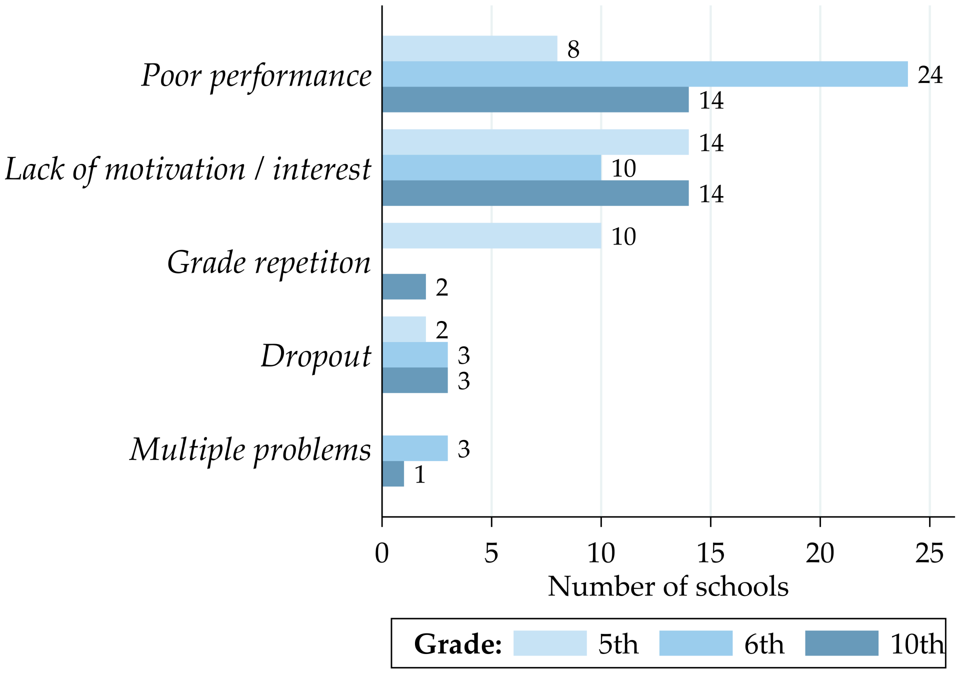
Online Appendix

A Pedagogical Projects

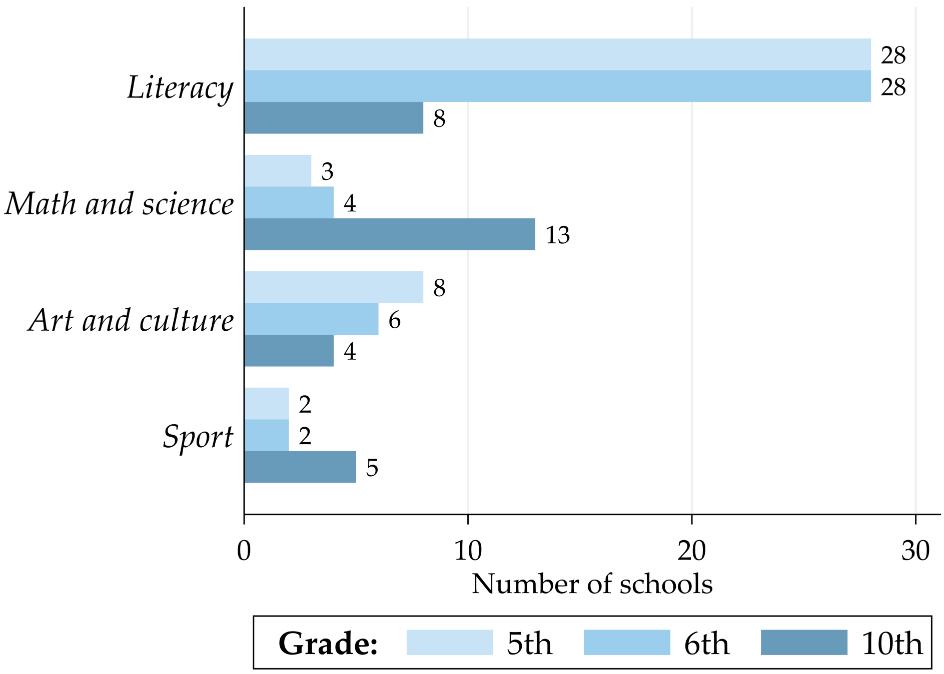
The Pedagogical Innovation Project (PIP) we study in this paper encouraged teachers to propose and introduce novel, locally tailored pedagogical projects in their classrooms. Hereafter, we provide some detailed examples of innovative pedagogical projects as planned and to be implemented in three schools in our experimental sample. These summaries are based on administrative information shared by experimental schools with the State Secretariat of Education.

Figure A1: Type of Pedagogical Projects by Grade

(a) Problems Addressed



(b) Activities Implemented



Note: The bars show the number of schools by main problem, as identified by teachers in their work plan for the PIP pedagogical activities, by type of activities implemented in the PIP projects, and by targeted grade. Data from State Secretariat of Education (SEE). Sample: schools in the treatment group.

Example 1: *Art and Culture*

Problem. Abandonment and evasion.

Contexts, cause and solution of the problem. Students attending this institution are from the community where the school is located as well as from adjacent municipalities. They come from a very heterogeneous socioeconomic context, including vulnerable segments of the population. Their parents engage in various autonomous income-generating activities such as agriculture, cassava processing, laundry services, and domestic work. Additionally, some families rely solely on the federal government's Bolsa Família social program for their income.

Given the lack of access to cultural and leisure goods, the school represents a reference point for knowledge acquisition and socialization with other peers. However, classes targeted by the PIP experienced a notable dropout rate in the previous year. The project team believes that the absence of aspirations is the root cause of high dropout. The future outlook for these young individuals is not consistently connected to their academic studies. On the other hand, the school itself has struggled to develop pedagogical practices that alleviate the ongoing challenges of student abandonment and evasion.

In light of the specific profile of these students, the project team opted to work on the theme of *Art and Culture*. They believe that engaging with art and culture through innovative methodologies can generate a bond between students, learning, and the school space itself. In addition, pedagogical initiatives will emphasize the development of reading and writing skills across various knowledge domains.

Main goal: Develop actions that encourage the school community to positively empower teachers and students through art and culture.

Specific objectives: Value artistic skills to promote learning in all disciplines; systematize activities that encourage student protagonism and genuine understanding of the content; provide space for interaction between different areas of knowledge; awaken artistic culture to value humanity; help young people open new paths to culture and the arts for their self-integration in the school system.

Target objectives: Ensure that at least 80% of students remain in the school by the end of the year; reduce dropout rates by 50%; increase passing rates by 40%.

Methodology and pedagogical activities. In order to meet the innovative needs of the PIP, the projects aim to create an environment room to subsidize innovative and interdisciplinary actions through arts and culture. This includes the organization of seminars, plays related to themes discussed in the classroom, and the execution of workshops throughout the academic year. Sound material, costumes and other accessories will be available in this room. Field classes will cover languages and human sciences, historical corridors in the city of Natal, museum visits, theater, and cinema.

The teaching team plans to: (i) form a permanent theater group with the aim of staging plays that rescue local culture with historical and para-folkloric reinterpretations; (ii) set up a permanent hip-hop dance group and a choir that covers classical music. In the field of visual arts, an exhibition by local artists and students' own productions will be organized for the entire school community. Finally, writing plots for theatrical plays and promoting soirées for the school community will address the need to enhance reading and writing skills.

Example 2: *Reading and Literacy*

Problem. Low performance.

Cause of the problem. The use of teaching methodologies that do not satisfactorily meet the needs of students generate demotivation and lack of interest among students.

Solution of the problem. The PIP is an important tool to support and encourage the insertion of new methodologies aimed at improving the teaching and learning process. Given the problems experienced by 4th and 5th year students at our school, there are noticeable difficulties in the process of acquiring reading, writing, and mathematical concepts, which is why we decided to choose the "Reading and Literacy" field for informing the project. More specifically, we aim at promoting the insertion and democratic use of technologies in the school environment in order to tackle to the issue of digital inclusion of children from the first years of schooling.

The school intends to innovate in its methodologies and seek not only to introduce new technologies into the classroom, but also to encourage a new attitude on the part of educators so that significant improvements in learning can occur. Students already use and dominate the language of technologies and media resources, such as smartphones, tablets, and computers. This activity is often completely disconnected to school teaching. Thus, the project's premise is to use playful methods and technological resources in order to offer a computerized educational environment, focused on the needs of students, and so make teaching more effective.

Main goal: Develop student reading, writing and numeracy skills in an interactive, innovative, enjoyable and playful way, aiming to reinforce and consolidate learning practices.

Specific objectives: encourage the creation of a real literacy environment through interaction between readers and readings and the insertion of ICTs in school activities; continuously monitor and revitalize teaching actions, projects and proposals, prioritizing real needs of the school institution; intensify the partnership between family and school, in the academic monitoring of students, making them aware of their responsibility for the success of learning; expand the possibilities of practicing literacy based on texts of social circulation that favor the enrichment of the repertoire cultural; feature photographic, camcorder, computers to register and communicate ideas; develop mastery of written expression as a personal word, becoming skilled in recording your ideas, opinions, feelings, memories; enlarge the diversity in textual genres known to children; guarantee a repertoire of good quality texts that constitute useful material consultation for writing other texts; understand the need to preserve school space and cultivate good coexistence to create a pleasant and fun environment; encourage the creation of a real literacy environment through the interaction between readers and readings; integrate schools and families through the use of the traveling suitcase, since these books will be read by the child and an adult in the family; use different types of mathematical languages and use them in your arguments; interpret and produce numerical writing through oral language, informal records and mathematical language, considering the rules of the decimal number system SND; develop an investigative, critical and creative spirit, in the context of problem situations, producing records own and seeking different solution strategies; carry out collective planning on a monthly basis with all segments of the company school to organize and coordinate actions to be taken, aiming to raise the level of student performance, involving them in social practices of reading, writing and numeracy.

Target objectives: ensure the participation of 100% of students in the 4th and 5th year of elementary school in reading, writing and numeracy; raise passing rates from 96% to at least 98%; guarantee the participation of 100% of teachers in the execution of the project; form a partnership between families and the school.

Methodology. To guarantee quality education, the school will seek to change its pedagogical practice by innovating and diversifying the teaching methodology in the classroom; concentrating efforts to adopt meaningful and effective methodologies, proposing diverse activities based on the students' reality, as well as seeking support from parents to overcome difficulties detected. Through diversified activities, the school seeks to create conditions for the formation of individuals capable of using information and not only assimilating content, but who can articulate knowledge, skills and values. Improving social interaction and striving for tolerance and freedom and helping the student to form concepts, build historical knowledge and act as a subject of their own learning.

Pedagogical Activities: reading, analysis and text production exploring different textual genres, including genres of the digital age; promotion of events that enable the presence and participation of students' families at school; pedagogical workshops dealing with the use of ICTs; mathematics championship; periodic studies to plan and monitor project results; literary picnic; exhibition of works carried out by students; using games to deepen knowledge and develop specific skills in the various curricular components; creating and recording parodies; listening and interpreting music of different styles; production of texts: graphics, opinion text, interview, letter, among others; use of the bibliographic collection for handling, reading and researching books; poetic soiree; edition of printed and digital newspapers; dynamic use of the digital whiteboard as a pedagogical tool; onstruction and analysis of graphs and tables, geometric figures, models, etc.; photographic recording of situations and/or environments, and writing of texts in different genres, based on photographic images; research using the internet as a search and consultation tool for school work; conducting interviews with people in the community on topics of social relevance; cultural sample with an exhibition of the work developed during the project; creation of a Facebook page to publicize PIP activities; reading competitions for various types of text, for example, tongue twisters, riddles, jokes, poems, narrative texts, etc.; reading wheel; video screening (films, documentaries and videos available on school TV); writing and mathematics workshops with students, using games and educational software in the computer laboratory.

Example 3: ***Scientific Initiation and Research***

Problem: Abandonment and evasion.

Cause of the problem: limitations of pedagogical practices; learning deficit; lack of family support and future perspectives among students.

Context and solution of the problem. The school serves a very diverse clientele, in terms of socioeconomic conditions, as it is the only high school in the city. Thus, students who complete elementary school in state, municipal, and private schools in the school municipality – and in some cases in neighboring municipalities – are enrolled in this teaching unit to complete their basic education.[49] The age range of the students, the level of learning and performance in different areas of knowledge are also quite heterogeneous,

[49]The classes served by PIP are made up of the afternoon shift (with students from rural areas, low education base, lacking skills in the use of technological resources) and the night shift (with students coming mainly from the local shipyard, working during the day – notably, many students abandoned their studies and are returning after a few years of work).

with the latter heterogeneity being particularly marked in reading and mathematics.

Given the school context, there exists an evident need for pedagogical practices that aim at: (i) minimizing the learning difficulties presented by students who have reached high school, with a particular emphasis on reading, research, and scientific investigation; (ii) enabling students to reorganize their ideas, create and recreate world views, fully exercise citizenship through education. The school has been constantly seeking to adapt to the evolving changes and transformations within the social environment. In pursuit of this adaptation, the school perceives the PIP as a viable and necessary alternative to leverage the restructuring process of the current teaching-learning framework, thus addressing the pressing issue of school dropout – a concern that weighs heavily on this educational institution.

Based on this assumption, the school chose "Scientific Initiation and Research" as its field of development, in that scientific methodology must be part of the secondary school student's curriculum, in order to make them the protagonist of their learning. Students can, in addition to satisfying their curiosities, develop critical thinking by establishing relationships between theory and practice and, therefore, repositioning themselves in the face of social, cultural, economic and political issues. In particular, we seek to include the use of Information and communications technology (ICT), reading, and mathematics, in every-day school life, as a means for pedagogical innovation.

Main goal: Combat school dropout by promoting access and ensuring the student's permanence in school; involve all the school community to secure broad attendance and learning.

Specific objectives: Use ICTs, such as Word, Excel, internet, emails, blogs, etc., as fundamental tools in the teaching-learning process; encourage reading, writing, and scientific reasoning in different areas of knowledge using the library space; support the student to be a protagonist in the learning process; promote activities that involve the family more consistently in life children's education; train teachers on the use of ICTs; propose and carry out actions that motivate the student to stay in school; search for more effective methods that help overcome students learning difficulties, awakening interest in research and scientific initiation.

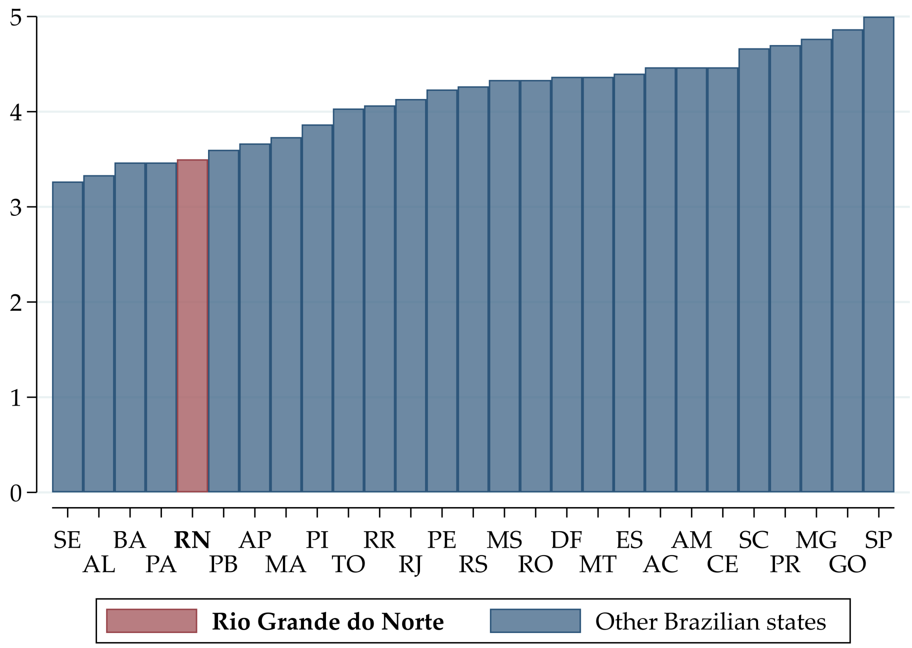
Target objectives: Reduce school dropout rates by 75% throughout the school year; involve 100% of students in educational activities, scientific initiation and research; train 85% of teachers, coordinators and pedagogical support in the use of ICTs, in the first two months; promote the practice of reading and writing on a daily basis; involve 60% of parents in activities developed at school during the school year.

Methodology and pedagogical activities. The teaching methodology adopted will seek to encourage the intellectual autonomy of both students and teachers. For this, different forms of teaching approaches, such as debate, observation, awareness raising, and use of web tools, will be leveraged. In addition, music and the production of documentary videos will provide an exercise for this autonomous expression. Spaces available at school, such as library and laboratories, will be used to enhance learning of students.

Activities to be carried out are the following ones: use of ICTs and other technological resources for the production and presentation of knowledge; reading and debate circles about various themes; practical classes using the science laboratory; bibliographic review on topics to be covered in scientific initiation projects; preparation of research questionnaires for data collection; assembly and analysis of graphs to disseminate the research results; use of graphs, probability and spatial geometry to emphasize the importance of mathematics in everyday life; creation of mathematical games and online challenges; creation of a Facebook page feed and a WhatsApp group with research, textual productions, and dissemination of school activities; organization of soirees, exhibitions, and workshops.

B Descriptive Statistics

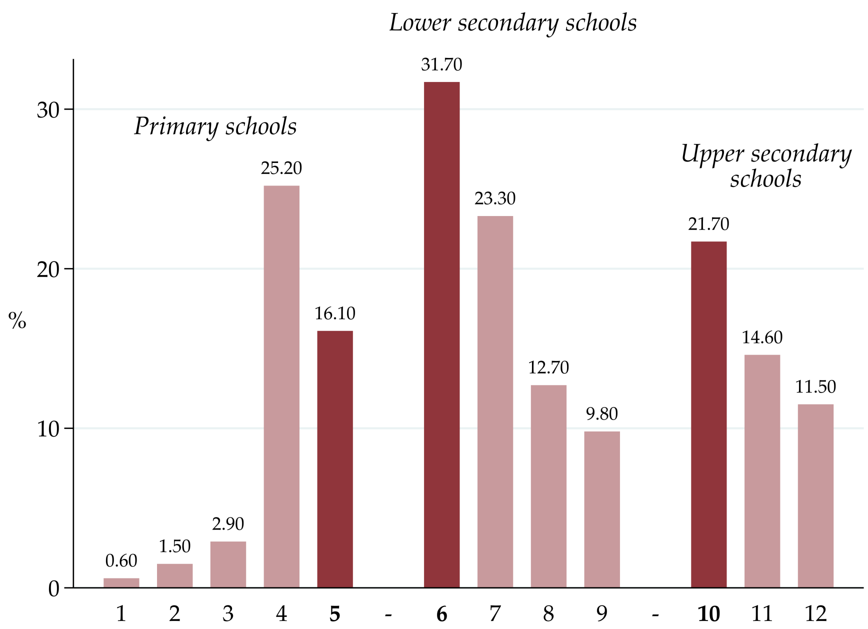
Figure B1: IDEB in Rio Grande do Norte vs. Other Brazilian States, 2015



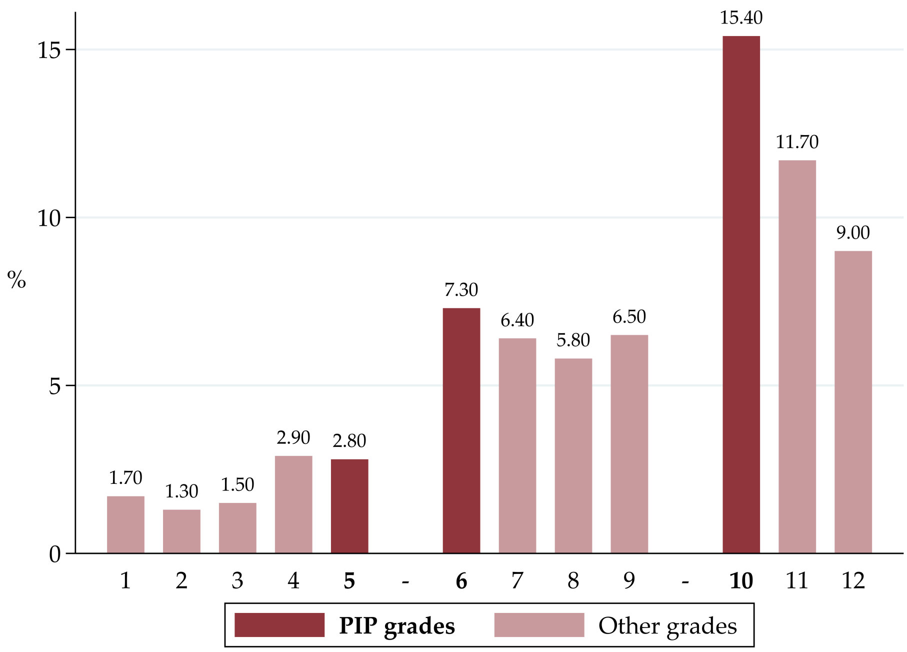
Note: We use data from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) for state public schools. The IDEB index is defined at each education stage, i.e., for primary, middle, and secondary schools. It is a national indicator for the quality of education, which combines information on student test scores and passing rates (see online Appendix F.1 for details on the construction of the index). The bars show the average IDEB across the three education stages by state in 2015.

Figure B2: Grade Repetition and School Dropout Rates in Rio Grande do Norte

(a) Grade Repetition Rate

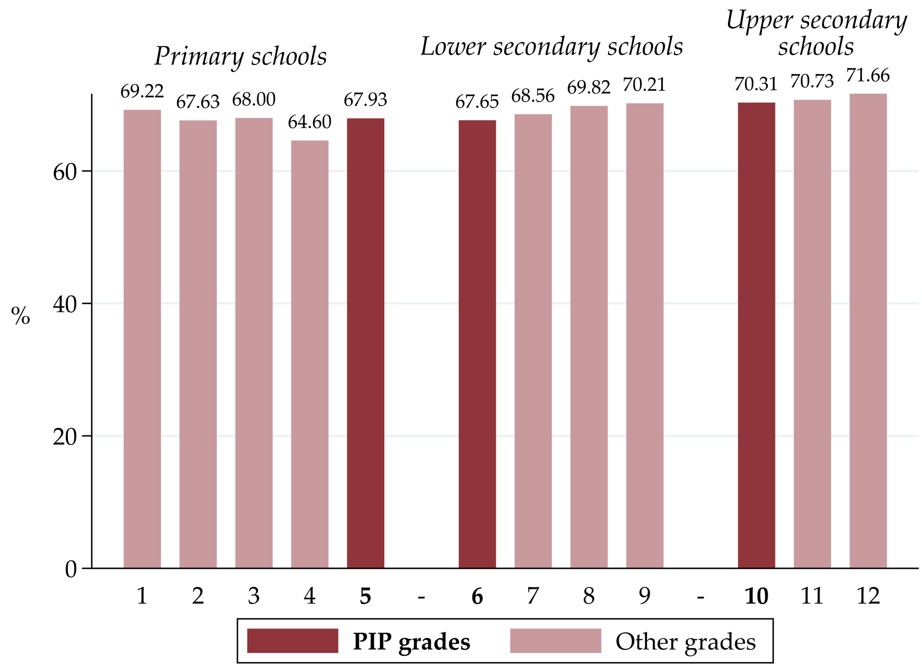


(b) School Dropout Rate



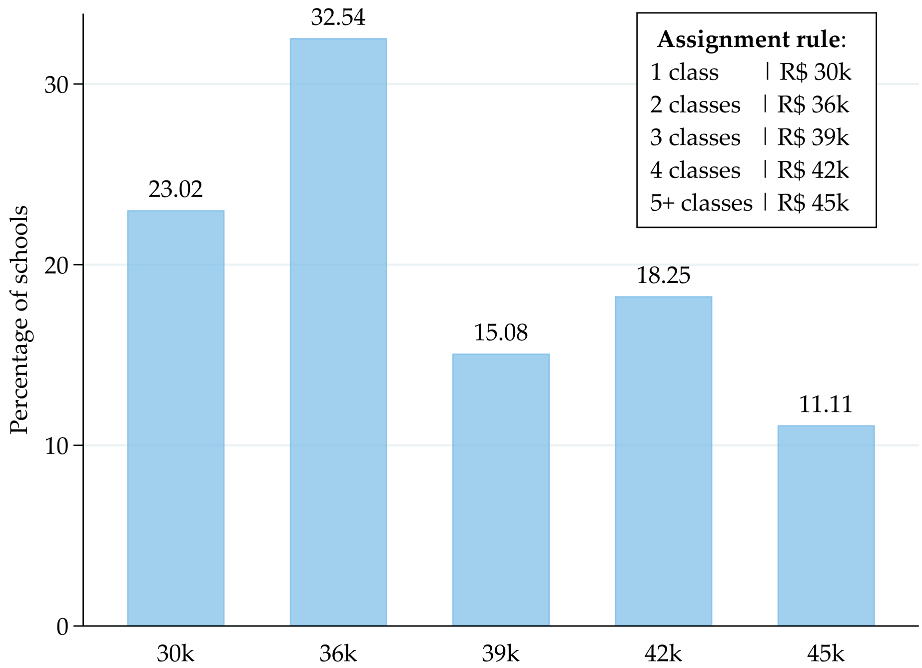
Note: Data are from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP). The bars show average repetition and dropout rate by grade among state public schools in Rio Grande do Norte in 2015.

Figure B3: Teacher Retention Rates in Rio Grande do Norte



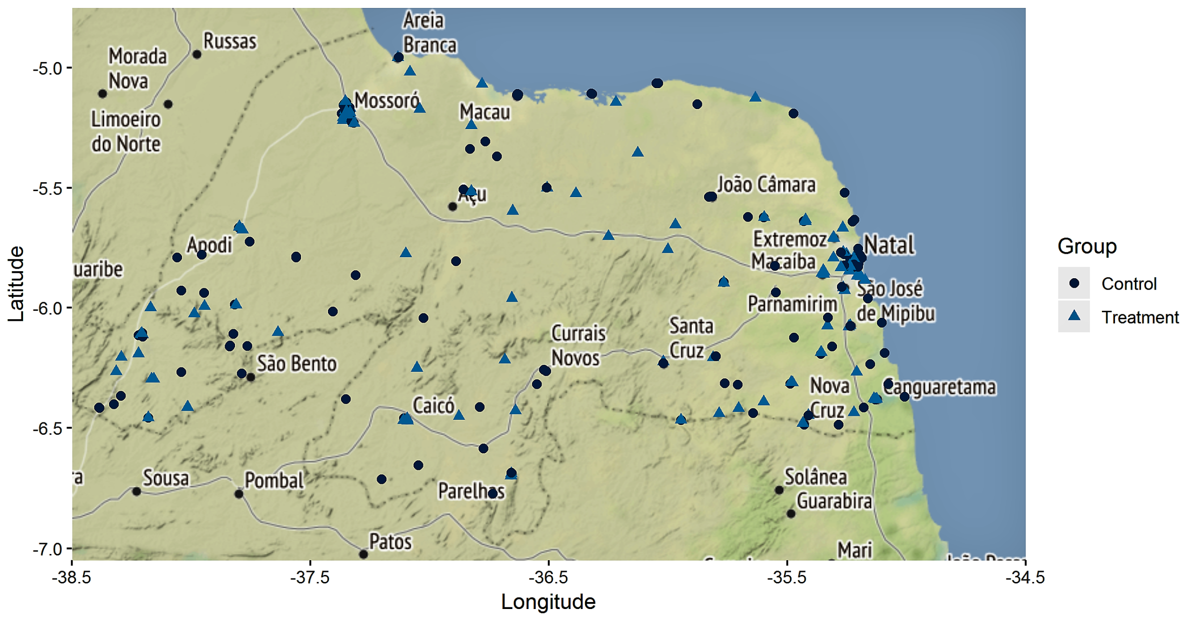
Note: Data are from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) teacher census. The bars show average retention rate by grade among state public schools in Rio Grande do Norte in 2015.

Figure B4: Allocation of Resources by Type of Grant



Note: The bars show the percentage of schools by the type of grant they were assigned to receive through PIP (ranging from 30,000 to 45,000). The values are in Brazilian *reais*, which were worth 0.25 US dollars at the beginning of 2016.

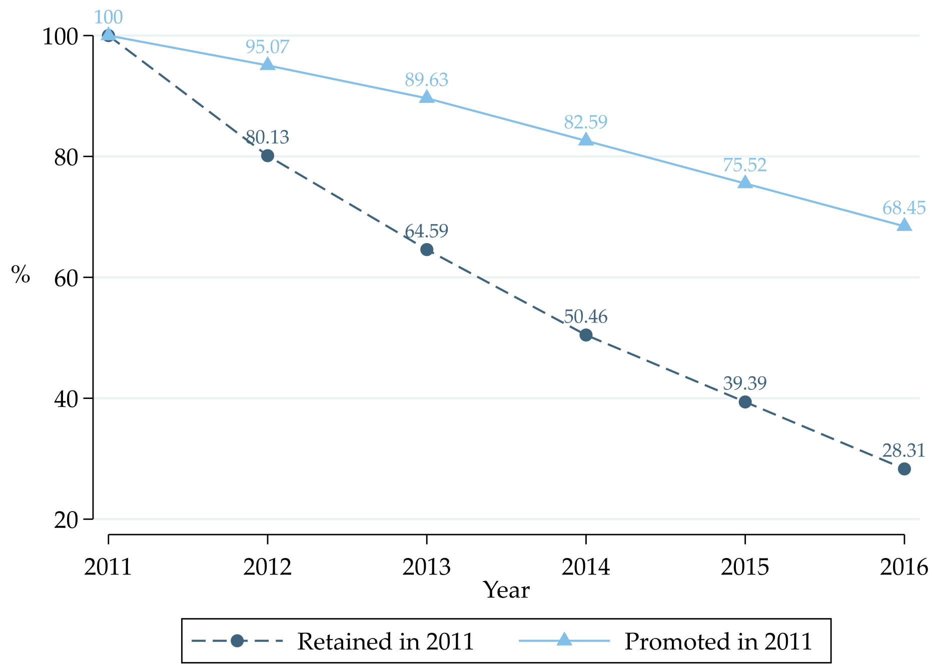
Figure B5: Geographical Distribution of Schools by Treatment Assignment



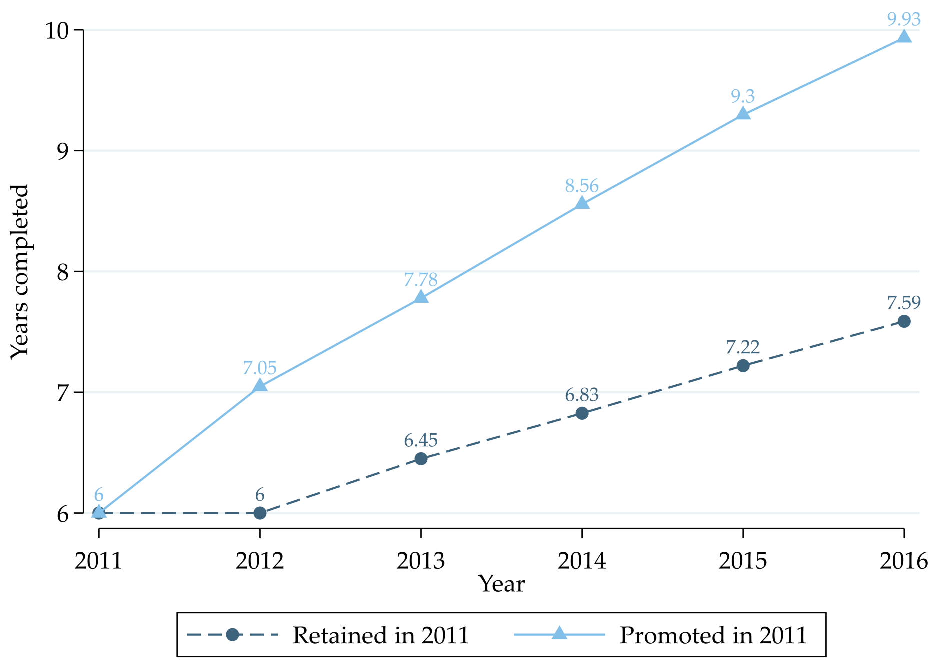
Note: GPS locations were extracted by scraping Google Maps API with school names. All but 6 schools in the experimental sample, 3 in the control and 3 in the treatment group, were not properly located using this method.

Figure B6: 6th Grade Repetition and Student Attainment

(a) Percentage of 2011 6th Graders Enrolled in Subsequent Years

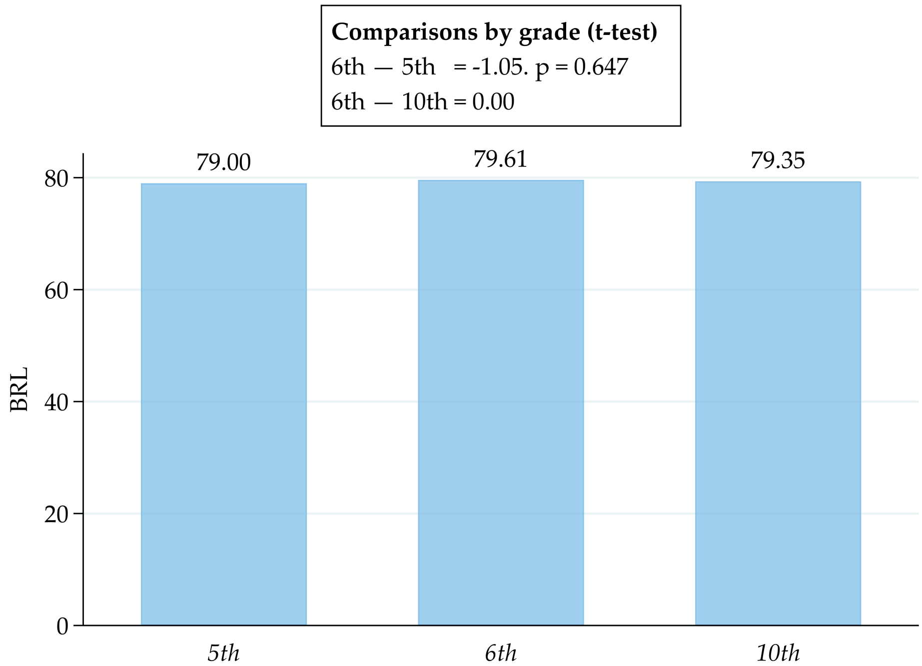


(b) Years of Completed Schooling of 2011 6th Graders



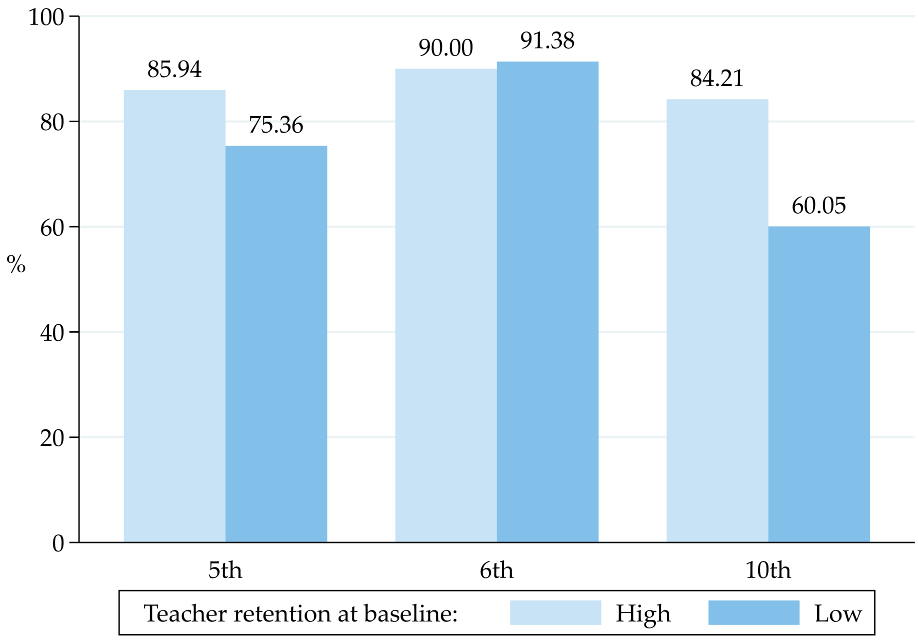
Note: The points in Panel (a) show the percentage of 6th graders in 2011 who were enrolled in any grade (6th or higher) in the following years, up to 2016. Panel (b) shows the average years of completed schooling of students who were enrolled in 6th grade in 2011 by each following year, up to 2016. The sample is the universe of students at public schools in Rio Grande do Norte (N = 73,010) and is split between those who were promoted in 2011 and those who were retained in 2011. Data from 2011-2017 school censuses.

Figure B7: Quality of Proposal by Grade



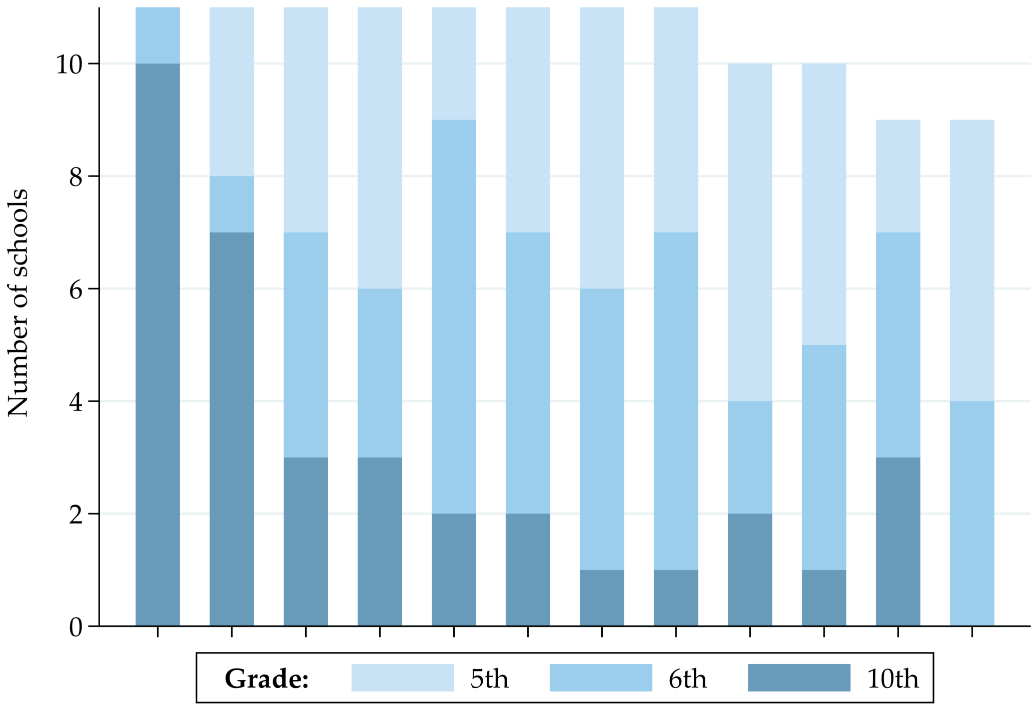
Note: The bars show the unconditional mean score of the expression of interest by targeted grade (range 0-100)). Comparisons by grade are done using region-fixed effects and robust standard errors. P-values based on standard t-test. Sample: schools in the treatment group. P-values based on standard t-test. Sample: schools in the treatment group.

Figure B8: Implementation by Teacher Retention at Baseline



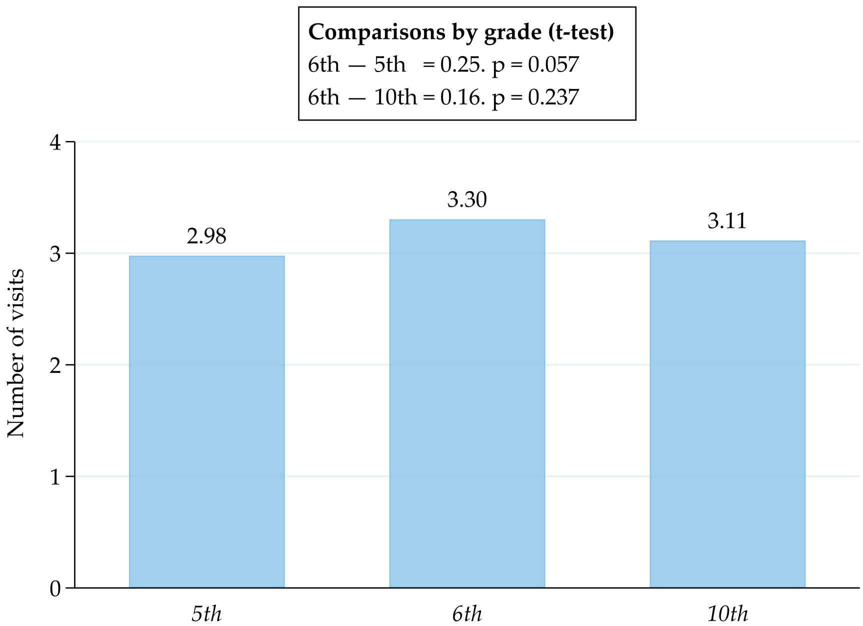
Note: 'Implementation' is defined as the ratio of the number of activities that were implemented over the number of planned activities described in the work plan. The bars show the unconditional mean of this variable by grade and baseline level of teacher retention. Data are from the State Secretariat of Education (SEE) and Rio Grande do Norte 2016 and 2017 teacher censuses. Sample: schools in the treatment group.

Figure B9: Mentors' Assignment by Grade



Note: The bars show the number of schools assigned to each mentor for supporting the development of the innovative pedagogical projects by grade. Each column indicates a different mentor. Sample: schools in the treatment group.

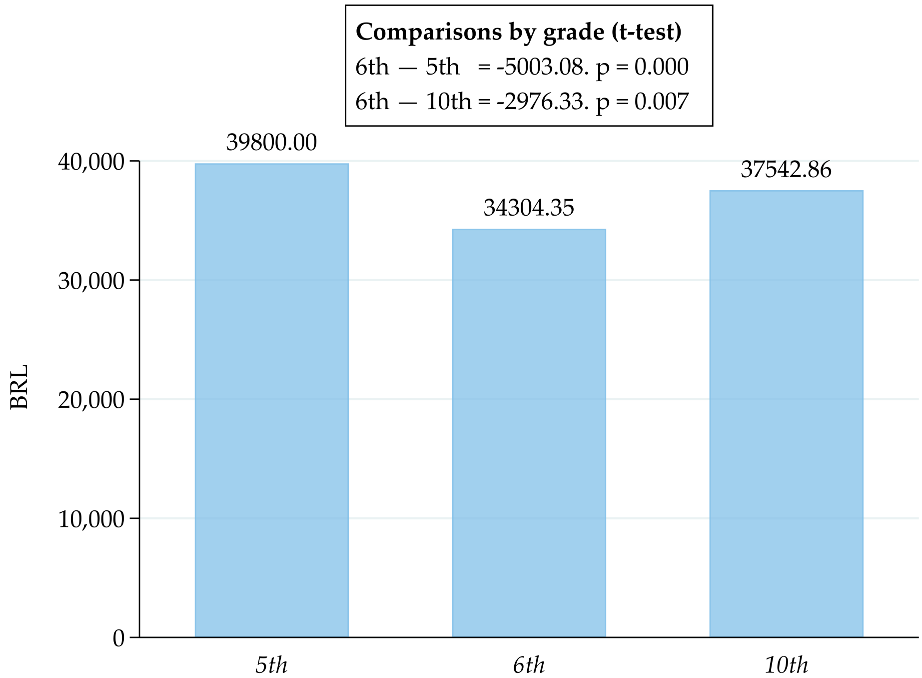
Figure B10: Mentors' Visits by Grade



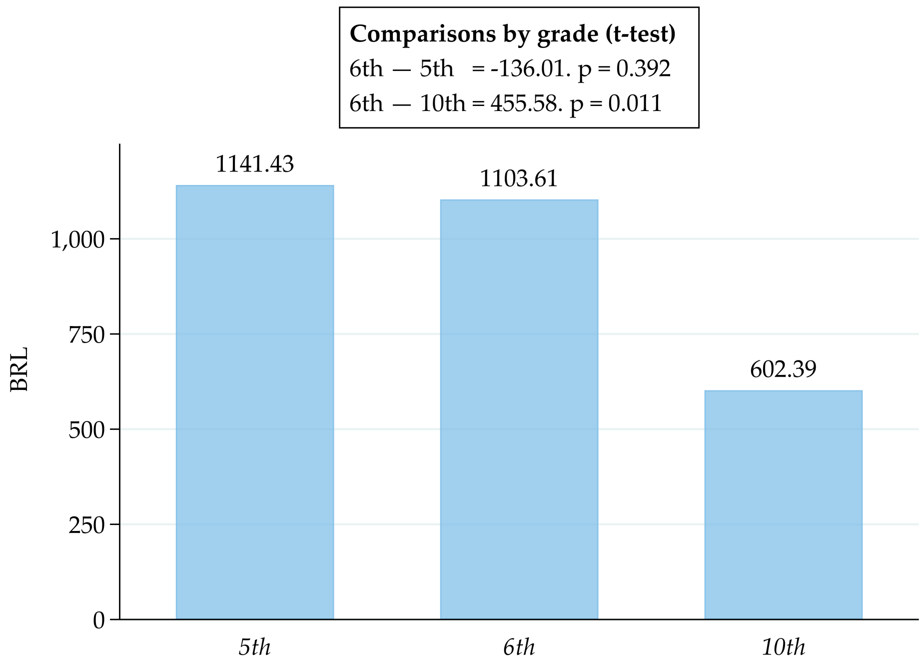
Note: The bars show the unconditional mean of the number of visits by targeted grade. Comparison by grade is done using Poisson regression with region fixed effects and robust standard errors. Differences between grades should be interpreted in terms of average marginal effect. Standard errors are then computed using the delta method and p-values based on standard t-test. Sample: schools in the treatment group.

Figure B11: Project Grant by Grade

(a) Value of the *Total* Grant



(b) Value of the Grant *per Student*



Note: The bars show the unconditional mean of the grant by targeted grade. The values are in Brazilian *reais*, which were worth 0.25 US dollars at the beginning of 2016. Comparisons by grade are done using region fixed effects and robust standard errors. *P*-values based on standard *t*-test. Sample: schools in the treatment group.

Table B1: Sample

A) Number of eligible schools			
	Treatment	Control	Total
5th grade	47	60	107
6th grade	48	63	111
10th grade	35	46	81
Total	130	169	299

B) Effective number of schools			
	Treatment	Control	Total
5th grade	45	52	97
6th grade	46	59	105
10th grade	35	43	78
Total	126	154	280

C) Number of enrolled students			
	Treatment	Control	Total
5th grade	4061	3952	8013
6th grade	2517	2871	5388
10th grade	2854	3644	6498
Total	9432	10467	19899

Table B2: Descriptive Statistics by Teacher Compliance

	Non-compliers		Compliers		
	(1)	(2)	(3)	(4)	(5)
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	Diff
Age	1546	40.103	302	40.705	-0.602
	246	(0.270)	92	(0.553)	
Gender (male = 1)	1546	0.493	302	0.493	-0.000
	246	(0.013)	92	(0.031)	
White	1073	0.501	203	0.547	-0.045
	234	(0.018)	73	(0.036)	
Has completed tertiary education	1546	0.937	302	0.950	-0.013
	246	(0.008)	92	(0.014)	
Has specialization and/or master	1546	0.393	302	0.437	-0.044
	246	(0.015)	92	(0.033)	
Teaches math	1538	0.218	302	0.242	-0.023
	246	(0.010)	92	(0.026)	
Teaches Portugues	1538	0.244	302	0.281	-0.037
	246	(0.011)	92	(0.026)	
Teaches science	1116	0.252	211	0.360	-0.108***
	215	(0.013)	79	(0.036)	
Teaches history	1538	0.223	302	0.291	-0.068**
	246	(0.011)	92	(0.030)	
Teaches geography	1538	0.223	302	0.248	-0.025
	246	(0.011)	92	(0.026)	
Teaches art	1538	0.254	302	0.252	0.003
	246	(0.013)	92	(0.024)	
Teaches phyisical education	1538	0.193	302	0.215	-0.022
	246	(0.011)	92	(0.022)	

Note: **Significant at 5 percent. ***Significant at 1 percent. For school and grade level comparisons we use data from the 2015 Rio Grande do Norte school census (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* – INEP) and progression rates from *Sistema Integrado de Gesto da Educao* (SIGEduc) portal. At the teacher and student level, we compare socio-demographics at the beginning of the year of the intervention, i.e., 2016, from that year Rio Grande do Norte school census. Teacher data regard only those teachers who taught in the classes involved in the program, and not from other grades. Standard errors (SE) clustered at the school level for the remaining rows in Panel B. Strata (i.e., region and grade) fixed effects are included in all the estimated regressions. p-values are computed using standard hypothesis testing.

Table B3: Descriptive Statistics – Comparison by Grade

Variable	5th grade (1) N/[Clusters]	5th grade (2) Mean/SE	6th grade (3) N/[Clusters]	6th grade (4) Mean/SE	10th grade (5) N/[Clusters]	10th grade (6) Mean/SE	5th - 6th (7) Diff	10th - 6th (8) Diff
Panel A – School characteristics								
Has access to internet	104	0.933	96	0.948	78	0.936	-0.015	-0.003
		(0.025)		(0.023)		(0.028)		
Has library	104	0.654	96	0.490	78	0.897	0.164**	-0.244
		(0.047)		(0.051)		(0.035)		
Has sciences lab	104	0.029	96	0.000	78	0.513	0.029	-0.484***
		(0.016)		(0.000)		(0.057)		
Located in urban area	105	1.210	97	1.082	78	1.154	0.127***	0.056***
		(0.040)		(0.028)		(0.041)		
Distance to Natal (km)	103	162.310	94	129.017	77	151.797	33.293***	10.514***
		(10.469)		(12.403)		(12.814)		
Number of employees	104	30.740	96	24.229	78	34.321	6.511	-3.580*
		(1.214)		(1.062)		(1.933)		
Number of students	104	377.404	96	271.250	78	473.026	106.154	-95.622***
		(19.120)		(16.006)		(40.566)		
Number of classes	104	15.423	96	12.281	78	16.449	3.142	-1.026*
		(0.644)		(0.745)		(1.323)		
Students per class	104	23.748	96	22.275	78	28.162	1.473	-4.414***
		(0.561)		(0.457)		(0.741)		
Panel B – Grades assigned to the intervention								
Passing rate	103	65.579	78	65.135	96	83.220	0.444	-17.641
		(1.653)		(1.811)		(1.331)		
Drop-out rate	103	6.994	78	17.210	96	1.868	-10.216***	5.126**
		(0.742)		(1.294)		(0.408)		
Retention rate	103	27.427	78	17.655	96	14.912	9.772	12.515
		(1.468)		(1.641)		(1.157)		
Teacher retention rate	98	0.648	78	0.692	63	0.828	-0.044	0.181**
		(0.020)		(0.022)		(0.031)		
Panel C – Teacher characteristics								
Age	784	40.417	909	40.021	189	40.164	0.396	0.253
	[104]	(0.349)	[78]	(0.386)	[95]	(0.639)		
Gender (male = 1)	784	0.462	909	0.567	189	0.233	-0.105***	0.229
	[104]	(0.016)	[78]	(0.016)	[95]	(0.034)		
White	545	0.508	616	0.502	136	0.537	0.007***	-0.029**
	[104]	(0.025)	[78]	(0.025)	[74]	(0.045)		
Has completed tertiary education	784	0.943	909	0.945	189	0.889	-0.002**	0.054**
	[104]	(0.010)	[78]	(0.010)	[95]	(0.026)		
Has specialization and/or master	784	0.423	909	0.384	189	0.360	0.040***	0.064**
	[104]	(0.020)	[78]	(0.021)	[95]	(0.035)		
Panel D – Student characteristics								
Age	5312	12.312	5124	16.247	7949	10.370	-3.935***	1.941***
	[104]	(0.070)	[66]	(0.066)	[96]	(0.040)		
Gender (male = 1)	5314	0.547	5124	0.457	7950	0.534	0.090***	0.013
	[104]	(0.009)	[66]	(0.008)	[96]	(0.007)		
White	3564	0.338	3593	0.320	4907	0.368	0.018***	-0.031***
	[101]	(0.028)	[65]	(0.025)	[94]	(0.026)		
Receives *Bolsa Famlia*	5314	0.344	5124	0.238	7950	0.348	0.106***	-0.004**
	[104]	(0.026)	[66]	(0.030)	[96]	(0.028)		

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. For school and grade level comparisons we use data from the 2015 Rio Grande do Norte school census (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* – INEP) and progression rates from *Sistema Integrado de Gesto da Educao* (SIGEduc) portal. At the teacher level, we compare socio-demographics at the beginning of the year of the intervention, i.e., 2016, from that year Rio Grande do Norte school census. Teacher data regard only those teachers who taught in the classes involved in the program, and not from other grades. Standard errors (SE) are robust in Panel A and the first three rows of B, and clustered at the school level for the remaining rows in Panel B. Strata (i.e., region and grade) fixed effects are included in all the estimated regressions. p-values are computed using standard hypothesis testing.

C Balance and Attrition

This section conducts a formal analysis of both differential and selective attrition in test-taking and checks whether balance is maintained in the sub-sample of test-takers. Based on these estimates, we draw implications for both internal and external validity of our experiment.

As we anticipated in Section 3.3, only a fraction of schools and students in the original study sample participated in the socio-emotional and proficiency test. Overall, 84 percent of schools participated in the socio-emotional test and 94 percent participated in the state standardized tests. Among the participating schools, on average, 55 percent of enrolled students took the socio-emotional test, and 69 percent participated in the proficiency tests.[50] Online Appendix Table C2 compares participation rates between the control and treatment group. Treated schools are more likely to participate in the socio-emotional test (91 versus 78 percent), while participation is balanced for the proficiency tests. Conditional on the school participating, the percentage of test takers is balanced for both tests, across all grades, suggesting no differential within school selection by treatment assignment. Note that we have imperfect overlap between students that took the socio-emotional test and those taking the proficiency test. Overall, 49 percent of students in the selected class took both tests, which restricts our ability to interact these variables in our analysis of the potential mechanism of the program.

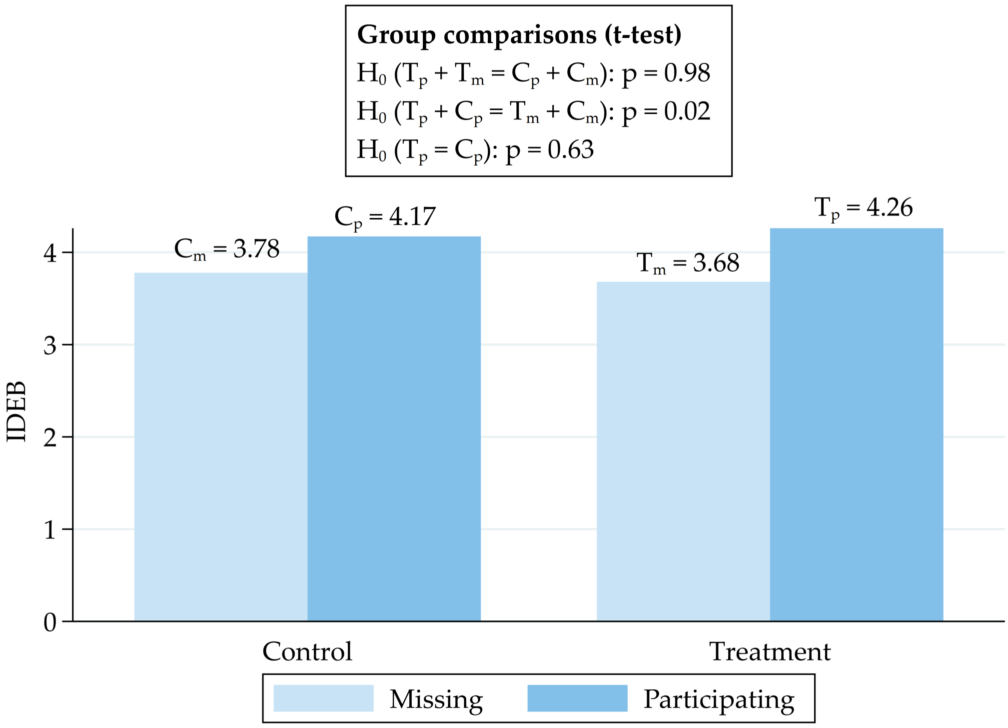
To explore how the unbalanced participation of schools in the socio-emotional test may affect our results, we replicate the balance table restricting the sample to schools with at least one test-taker (online Appendix Tables C3-C4). We find similar balance results between treatment and control schools among this sub-sample of test-takers, suggesting that the internal validity holds for the respondent sub-population. To test whether school quality varies across test-takers and non-test-takers, we compare schools that participated

[50]Lower participation in the socio-emotional test is explained by the fact that it was carried out later than the proficiency test, when some of the schools in our sample had already released their students for the summer break.

in the socio-emotional test with those who did not, across treatment and control groups (online Appendix Figure C1), using the 2015 IDEB as a measure of school quality at baseline. As expected, treatment and control schools have similar IDEB scores ($p = 0.98$). However, participating schools have better scores than non-participating schools ($p = 0.02$). Yet this pattern appears to be no different among treatment and control groups ($p = 0.62$).

We further characterize student test compliance in participating schools: online Appendix Table C5 shows that socio-emotional test-takers are, on average, younger across treated grades, while there are no statistically significant differences in the other characteristics we have data on. On the other hand, there are some grade-specific imbalances when considering proficiency test-takers.

Figure C1: IDEB by School Participation to Socio-Emotional Test and Treatment



Note: The bars show the unconditional means of the school IDEB by participation in the socio-emotional test and by treatment assignment, as described in 3.3. We regress IDEB on these 4 categories so that:

$$IDEB_s = \beta_1 \cdot T_{m_s} + \beta_2 \cdot T_{p_s} + \beta_3 \cdot C_{m_s} + \beta_4 \cdot C_{p_s} + \varepsilon_s$$

Therefore, we run three different group comparisons – namely treated schools vs. control; participating schools vs. missing schools; treated vs. control among participating schools – by testing the null hypotheses that $\beta_2 + \beta_4 = \beta_1 + \beta_3$, $\beta_2 + \beta_1 = \beta_4 + \beta_2$, $\beta_2 = \beta_4$, respectively, through standard t-tests. IDEB data refer to 2015 and are from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP).

Table C1: Balance Table

	All schools					5th Grade	6th Grade	10th Grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					T-test	T-test	T-test	T-test
		Control		Treatment	P-value	P-value	P-value	P-value
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	[RI p-value]	[RI p-value]	[RI p-value]	[RI p-value]
Panel A – School characteristics								
Has access to internet	154	0.922	124	0.960	0.197	0.228	0.413	0.835
		(0.022)		(0.018)	[0.216]	[0.364]	[0.457]	[1.000]
Has library	154	0.669	124	0.661	0.957	0.970	0.301	0.046
		(0.038)		(0.043)	[0.963]	[1.000]	[0.334]	[0.070]
Has sciences lab	154	0.143	124	0.169	0.412	N/A	0.721	0.311
		(0.028)		(0.034)	[0.427]	[1.000]	[1.000]	[0.358]
Located in urban area	154	1.169	126	1.127	0.339	0.197	0.835	0.389
		(0.030)		(0.030)	[0.395]	[0.278]	[1.000]	[0.547]
Distance to Natal (km)	151	152.086	123	142.837	0.877	0.760	0.915	0.599
		(9.140)		(10.375)	[0.879]	[0.759]	[0.919]	[0.605]
Number of employees	154	29.422	124	29.589	0.889	0.824	0.718	0.510
		(1.136)		(1.222)	[0.889]	[0.830]	[0.709]	[0.509]
Number of students	154	361.903	124	374.621	0.697	0.636	0.904	0.854
		(19.116)		(24.671)	[0.686]	[0.631]	[0.901]	[0.852]
Number of classes	154	14.669	124	14.573	0.870	0.777	0.634	0.774
		(0.666)		(0.825)	[0.861]	[0.770]	[0.617]	[0.775]
Students per class	154	24.216	124	24.802	0.388	0.263	0.171	0.310
		(0.509)		(0.514)	[0.382]	[0.258]	[0.174]	[0.318]
Panel B – Grades assigned to the intervention								
Passing rate	154	70.714	123	72.637	0.389	0.372	0.194	0.412
		(1.434)		(1.545)	[0.387]	[0.372]	[0.188]	[0.416]
Drop-out rate	154	8.023	123	8.183	0.811	0.381	0.243	0.123
		(0.780)		(0.943)	[0.813]	[0.368]	[0.266]	[0.132]
Retention rate	154	21.263	123	19.180	0.265	0.475	0.372	0.761
		(1.201)		(1.298)	[0.262]	[0.481]	[0.360]	[0.761]
Teacher retention rate	130	0.700	109	0.721	0.606	0.903	0.977	0.239
		(0.021)		(0.020)	[0.612]	[0.903]	[0.976]	[0.257]
Panel C – Teacher characteristics								
Age	1021	40.296	861	40.087	0.666	0.213	0.744	0.704
	[153]	(0.331)	[124]	(0.363)	[0.630]	[0.235]	[0.724]	[0.637]
Gender (male = 1)	1021	0.471	861	0.511	0.229	0.884	0.123	0.687
	[153]	(0.016)	[124]	(0.019)	[0.292]	[1.000]	[0.197]	[0.731]
White	715	0.491	582	0.529	0.256	0.937	0.719	0.230
	[146]	(0.023)	[110]	(0.023)	[0.203]	[0.933]	[0.700]	[0.154]
Has completed tertiary education	1021	0.937	861	0.940	0.840	0.114	0.908	0.229
	[153]	(0.010)	[124]	(0.010)	[0.849]	[0.100]	[1.000]	[0.139]
Has specialization and/or master	1021	0.405	861	0.389	0.724	0.750	0.147	0.092
	[153]	(0.019)	[124]	(0.019)	[0.706]	[0.757]	[0.108]	[0.047]
Panel D – Student characteristics								
Age	9558	12.725	8827	12.401	0.088	0.275	0.059	0.987
	[146]	(0.266)	[120]	(0.276)	[0.091]	[0.285]	[0.075]	[0.990]
Gender (male = 1)	9560	0.517	8828	0.516	0.579	0.789	0.189	0.333
	[146]	(0.007)	[120]	(0.008)	[0.590]	[0.790]	[0.205]	[0.338]
White	6245	0.354	5819	0.335	0.244	0.341	0.660	0.566
	[142]	(0.020)	[118]	(0.024)	[0.288]	[0.362]	[0.696]	[0.629]
Receives *Bolsa Famlia*	9560	0.319	8828	0.313	0.947	0.496	0.262	0.860
	[146]	(0.025)	[120]	(0.024)	[0.949]	[0.498]	[0.288]	[0.877]

Note: For school and grade level comparisons we use data from the 2015 Rio Grande do Norte school census (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* – INEP) and progression rates from *Sistema Integrado de Gesto da Educao* (SIGEduc) portal. At the teacher and student level, we compare socio-demographics at the beginning of the year of the intervention, i.e., 2016, from that year Rio Grande do Norte school census. Teacher data regard only those teachers who taught in the classes involved in the program, and not from other grades. Student data regard students enrolled in those grades at the beginning of the school year. Two schools out of the 280 schools in the sample are missing in the census. Standard errors (SE) are robust in Panel A and B, and clustered at the school level in Panel C and D. Strata (i.e., region and grade) fixed effects are included in all the estimated regressions. We show both standard p-values and p-values computed using randomization inference (RI) with 10,000 repetitions for the whole sample and for each grade. The coefficient on 'Has science lab' in 5th grade is not available (N/A) in the table as there is no variation in such variable across experimental arms (namely, no 5th-grade school had a science lab in 2015, neither in the control nor in the treatment group).

Table C2: Balance in Socio-Emotional and Proficiency Test Participation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable		Total		Control		Treatment	T-test	RI
Sample	N	Mean/SE	N	Mean/SE	N	Mean/SE	P-value	P-value
Panel A – Socio-emotional tests								
Participating schools								
All schools	280	0.839	154	0.779	126	0.913	0.002	0.003
		(0.022)		(0.034)		(0.025)		
5th grade	97	0.876	52	0.827	45	0.933	0.125	0.151
		(0.034)		(0.053)		(0.038)		
6th grade	105	0.829	59	0.780	46	0.891	0.124	0.149
		(0.037)		(0.054)		(0.046)		
10th grade	78	0.808	43	0.721	35	0.914	0.025	0.038
		(0.045)		(0.069)		(0.048)		
Percentage of test takers								
All schools	235	0.549	120	0.530	115	0.570	0.180	0.184
		(0.015)		(0.021)		(0.021)		
5th grade	85	0.578	43	0.547	42	0.610	0.209	0.210
		(0.024)		(0.030)		(0.036)		
6th grade	87	0.545	46	0.539	41	0.551	0.823	0.826
		(0.024)		(0.034)		(0.033)		
10th grade	63	0.517	31	0.492	32	0.541	0.392	0.412
		(0.031)		(0.048)		(0.042)		
Panel B – Proficiency tests								
Participating schools								
All schools	280	0.943	154	0.942	126	0.944	0.941	1.000
		(0.014)		(0.019)		(0.020)		
5th grade	97	0.948	52	0.942	45	0.956	0.888	0.906
		(0.023)		(0.033)		(0.031)		
6th grade	105	0.943	59	0.949	46	0.935	0.698	0.688
		(0.023)		(0.029)		(0.037)		
10th grade	78	0.936	43	0.930	35	0.943	0.289	0.467
		(0.028)		(0.039)		(0.040)		
Percentage of test takers								
All schools	264	0.696	145	0.699	119	0.691	0.775	0.778
		(0.015)		(0.019)		(0.023)		
5th grade	92	0.408	49	0.418	43	0.396	0.256	0.264
		(0.008)		(0.009)		(0.013)		
6th grade	99	0.853	56	0.840	43	0.870	0.245	0.264
		(0.013)		(0.018)		(0.019)		
10th grade	73	0.845	40	0.848	33	0.841	0.720	0.723
		(0.016)		(0.022)		(0.025)		

Note: '*Participating schools*' is a dummy for schools that had at least one test taker. '*Percentage of test takers*' is defined as the percentage of students who took the test for each school in the sample, conditional on the school being a 'participating school'. Robust standard errors (SE) are in parentheses. Strata (i.e., region) fixed effects are included in all the estimated regressions. We show both standard p-values and p-values computed using randomization inference (RI) with 10,000 repetitions for the whole sample and each grade.

Table C3: Balance Table on Subsample of Schools with Socio-Emotional Test Takers

Variable	All schools (1) N/[Clusters]	(2) Control Mean/SE	(3) N/[Clusters]	(4) Treatment Mean/SE	(5) T-test P-value [RI p-value]	5th Grade (6) T-test P-value [RI p-value]	6th Grade (7) T-test P-value [RI p-value]	10th Grade (8) T-test P-value [RI p-value]
Panel A – School characteristics								
Has access to internet	120	0.917	115	0.957	0.227	0.182	0.501	0.947
		(0.025)		(0.019)	[0.288]	[0.373]	[0.680]	[1.000]
Has library	120	0.642	115	0.661	0.839	0.853	0.602	0.179
		(0.044)		(0.044)	[0.879]	[1.000]	[0.662]	[0.359]
Has sciences lab	120	0.125	115	0.174	0.243	N/A	0.636	0.137
		(0.030)		(0.035)	[0.254]	[1.000]	[1.000]	[0.203]
Located in urban area	120	1.175	115	1.113	0.179	0.173	0.747	0.438
		(0.035)		(0.030)	[0.249]	[0.267]	[0.783]	[0.500]
Distance to Natal (km)	117	143.815	114	148.243	0.739	0.895	0.681	0.732
		(10.416)		(10.806)	[0.742]	[0.896]	[0.678]	[0.736]
Number of employees	120	30.100	115	28.904	0.365	0.609	0.502	0.691
		(1.269)		(1.161)	[0.366]	[0.617]	[0.499]	[0.688]
Number of students	120	379.800	115	360.304	0.419	0.956	0.607	0.449
		(22.390)		(24.342)	[0.430]	[0.957]	[0.602]	[0.462]
Number of classes	120	15.425	115	14.096	0.160	0.600	0.234	0.426
		(0.787)		(0.815)	[0.164]	[0.605]	[0.230]	[0.442]
Students per class	120	24.107	115	24.721	0.352	0.325	0.275	0.587
		(0.572)		(0.540)	[0.352]	[0.326]	[0.268]	[0.581]
Panel B – Grades assigned to the intervention								
Passing rate	120	71.697	114	73.711	0.333	0.511	0.277	0.885
		(1.630)		(1.600)	[0.327]	[0.513]	[0.267]	[0.885]
Drop-out rate	120	7.428	114	8.026	0.715	0.355	0.869	0.332
		(0.848)		(0.953)	[0.717]	[0.360]	[0.869]	[0.324]
Repetition rate	120	20.876	114	18.263	0.195	0.668	0.249	0.567
		(1.371)		(1.316)	[0.186]	[0.671]	[0.248]	[0.559]
Panel C – Teacher characteristics								
Age	783	40.553	780	40.010	0.343	0.342	0.920	0.320
	[119]	(0.381)	[115]	(0.392)	[0.364]	[0.340]	[0.920]	[0.337]
Gender (male = 1)	783	0.466	780	0.513	0.096	0.927	0.086	0.381
	[119]	(0.019)	[115]	(0.020)	[0.120]	[0.929]	[0.098]	[0.418]
White	548	0.487	528	0.532	0.215	0.867	0.357	0.332
	[112]	(0.027)	[101]	(0.024)	[0.235]	[0.869]	[0.380]	[0.338]
Has completed tertiary education	783	0.948	780	0.941	0.545	0.063	0.862	0.657
	[119]	(0.009)	[115]	(0.010)	[0.552]	[0.060]	[0.867]	[0.660]
Has specialization and/or master	783	0.405	780	0.400	0.889	0.903	0.123	0.115
	[119]	(0.022)	[115]	(0.020)	[0.893]	[0.898]	[0.135]	[0.124]
Panel D – Student characteristics								
Age	7656	12.592	8011	12.407	0.091	0.224	0.228	0.608
	[116]	(0.294)	[112]	(0.292)	[0.100]	[0.238]	[0.259]	[0.618]
Gender (male = 1)	7657	0.515	8012	0.516	0.847	0.761	0.526	0.345
	[116]	(0.008)	[112]	(0.008)	[0.859]	[0.761]	[0.534]	[0.389]
White	4903	0.341	5405	0.341	0.403	0.332	0.763	0.942
	[113]	(0.023)	[110]	(0.026)	[0.427]	[0.366]	[0.786]	[0.951]
Receives *Bolsa Famlia*	7657	0.303	8012	0.306	0.964	0.665	0.412	0.790
	[116]	(0.026)	[112]	(0.025)	[0.964]	[0.672]	[0.423]	[0.779]

Note: For school and grade level comparisons we use data from the 2015 Rio Grande do Norte school census (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* – INEP) and progression rates from *Sistema Integrado de Gesto da Educao* (SIGEduc) portal. At the teacher and student level, we compare socio-demographics at the beginning of the year of the intervention, i.e., 2016, from that year Rio Grande do Norte school census. Teacher data regard only those teachers who taught in the classes involved in the program, and not from other grades. Student data regard students enrolled in those grades at the beginning of the school year. The sample is restricted to schools that had at least one socio-emotional test taker. Standard errors (SE) are robust in Panel A and B, and clustered at the school level in Panel C and D. Strata (i.e., region and grade) fixed effects are included in all the estimated regressions. We show both standard p-values and p-values computed using randomization inference (RI) with 10,000 repetitions for the whole sample and for each grade. The coefficient on 'Has science lab' in 5th grade is not available (N/A) in the table as there is no variation in such variable across experimental arms (namely, no 5th-grade school had a science lab in 2015, neither in the control nor in the treatment group).

Table C4: Balance Table on Subsample of Schools with Proficiency Test Takers

	All schools					5th Grade	6th Grade	10th Grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	N/[Clusters]	Control Mean/SE	N/[Clusters]	Treatment Mean/SE	T-test P-value [RI p-value]	T-test P-value [RI p-value]	T-test P-value [RI p-value]	T-test P-value [RI p-value]
Panel A – School characteristics								
Has access to internet	145	0.917	119	0.958	0.146	0.124	0.401	0.835
		(0.023)		(0.018)	[0.205]	[0.201]	[0.469]	[1.000]
Has library	145	0.669	119	0.647	0.821	0.822	0.224	0.046
		(0.039)		(0.044)	[0.893]	[0.839]	[0.285]	[0.071]
Has sciences lab	145	0.145	119	0.160	0.530	N/A	0.722	0.415
		(0.029)		(0.034)	[0.650]	[1.000]	[1.000]	[0.465]
Located in urban area	145	1.159	119	1.118	0.378	0.219	0.929	0.563
		(0.030)		(0.030)	[0.463]	[0.279]	[1.000]	[0.749]
Distance to Natal (km)	143	150.530	118	141.074	0.966	0.627	0.987	0.625
		(9.630)		(10.785)	[0.967]	[0.623]	[0.988]	[0.616]
Number of employees	145	29.669	119	29.462	0.984	0.862	0.682	0.666
		(1.194)		(1.245)	[0.984]	[0.868]	[0.678]	[0.670]
Number of students	145	367.324	119	377.429	0.700	0.695	0.831	0.872
		(19.942)		(25.575)	[0.699]	[0.700]	[0.828]	[0.866]
Number of classes	145	14.814	119	14.647	0.895	0.744	0.700	0.766
		(0.700)		(0.858)	[0.900]	[0.747]	[0.688]	[0.767]
Students per class	145	24.401	119	24.867	0.439	0.284	0.148	0.222
		(0.523)		(0.517)	[0.440]	[0.299]	[0.154]	[0.228]
Panel B – Grades assigned to the intervention								
Passing rate	145	69.765	118	72.897	0.147	0.279	0.065	0.599
		(1.459)		(1.551)	[0.146]	[0.301]	[0.062]	[0.594]
Drop-out rate	145	8.179	118	8.051	0.919	0.325	0.083	0.123
		(0.804)		(0.966)	[0.917]	[0.316]	[0.094]	[0.129]
Repetition rate	145	22.057	118	19.052	0.112	0.380	0.215	0.543
		(1.224)		(1.291)	[0.109]	[0.400]	[0.216]	[0.545]
Panel C – Teacher characteristics								
Age	973	40.276	815	40.124	0.838	0.275	0.817	0.952
	[144]	(0.343)	[119]	(0.376)	[0.847]	[0.258]	[0.820]	[0.952]
Gender (male = 1)	973	0.479	815	0.514	0.288	0.741	0.212	0.586
	[144]	(0.016)	[119]	(0.020)	[0.291]	[0.741]	[0.224]	[0.595]
White	675	0.474	543	0.523	0.150	0.869	0.475	0.202
	[138]	(0.023)	[105]	(0.024)	[0.164]	[0.876]	[0.517]	[0.200]
Has completed tertiary education	973	0.935	815	0.937	0.737	0.140	0.942	0.198
	[144]	(0.010)	[119]	(0.010)	[0.742]	[0.146]	[0.948]	[0.221]
Has specialization and/or master	973	0.400	815	0.385	0.882	0.811	0.140	0.140
	[144]	(0.020)	[119]	(0.020)	[0.885]	[0.799]	[0.156]	[0.151]
Panel D – Student characteristics								
Age	9201	12.770	8518	12.333	0.050	0.209	0.024	0.954
	[137]	(0.273)	[115]	(0.278)	[0.063]	[0.261]	[0.037]	[0.958]
Gender (male = 1)	9203	0.515	8519	0.515	0.669	0.947	0.169	0.278
	[137]	(0.008)	[115]	(0.008)	[0.679]	[0.949]	[0.187]	[0.287]
White	6025	0.337	5552	0.318	0.343	0.493	0.662	0.614
	[134]	(0.018)	[114]	(0.023)	[0.373]	[0.510]	[0.686]	[0.670]
Receives *Bolsa Famlia*	9203	0.317	8519	0.311	0.959	0.524	0.262	0.828
	[137]	(0.025)	[115]	(0.024)	[0.959]	[0.525]	[0.275]	[0.856]

Note: For school and grade level comparisons we use data from the 2015 Rio Grande do Norte school census (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* – INEP) and progression rates from *Sistema Integrado de Gesto da Educao* (SIGEduc) portal. At the teacher and student level, we compare socio-demographics at the beginning of the year of the intervention, i.e., 2016, from that year Rio Grande do Norte school census. Teacher data regard only those teachers who taught in the classes involved in the program, and not from other grades. Student data regard students enrolled in those grades at the beginning of the school year. The sample is restricted to schools that had at least one proficiency test taker. Standard errors (SE) are robust in Panel A and B and clustered at the school level in Panel C and D. Strata (i.e., region and grade) fixed effects are included in all the estimated regressions. We show both standard p-values and p-values computed using randomization inference (RI) with 10,000 repetitions for the whole sample and for each grade. The coefficient on 'Has science lab' in 5th grade is not available (N/A) in the table as there is no variation in such variable across experimental arms (namely, no 5th-grade school had a science lab in 2015, neither in the control nor in the treatment group).

Table C5: Comparison between Test-Taking and Non-Test-Taking Students

	All schools					5th Grade	6th Grade	10th Grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	N/[Clusters]	Test-takers Mean/SE	N/[Clusters]	Non-test-takers Mean/SE	T-test P-value [RI p-value]	T-test P-value [RI p-value]	T-test P-value [RI p-value]	T-test P-value [RI p-value]
Panel A – Socio-emotional tests								
Age	3419	12.595	2468	13.300	<0.000	<0.000	<0.000	<0.000
	[224]	(0.169)	[224]	(0.191)	[<0.000]	[<0.000]	[<0.000]	[<0.000]
Gender (male = 1)	3419	0.510	2468	0.524	0.238	0.557	0.340	0.662
	[224]	(0.010)	[224]	(0.011)	[0.246]	[0.586]	[0.334]	[0.638]
White	2141	0.362	1548	0.353	0.547	0.277	0.772	0.840
	[213]	(0.020)	[206]	(0.021)	[0.499]	[0.206]	[0.748]	[0.822]
Pardo	2141	0.610	1548	0.618	0.506	0.329	0.780	0.873
	[213]	(0.019)	[206]	(0.021)	[0.455]	[0.268]	[0.767]	[0.851]
Black	2141	0.025	1548	0.026	0.935	0.740	0.364	0.428
	[213]	(0.005)	[206]	(0.005)	[0.932]	[0.813]	[0.325]	[0.323]
Receives *Bolsa Famlia*	3419	0.338	2468	0.318	0.293	0.736	0.524	0.351
	[224]	(0.020)	[224]	(0.020)	[0.163]	[0.639]	[0.443]	[0.164]
Receives school transportation	3419	0.112	2468	0.136	0.062	0.327	0.160	0.367
	[224]	(0.012)	[224]	(0.015)	[0.012]	[0.148]	[0.101]	[0.162]
Panel B – Proficiency tests								
Age	11282	13.064	6437	11.676	0.475	<0.000	<0.000	<0.000
	[252]	(0.182)	[246]	(0.221)	[0.023]	[<0.000]	[<0.000]	[<0.000]
Gender (male = 1)	11285	0.504	6437	0.535	0.007	0.003	0.844	0.532
	[252]	(0.007)	[246]	(0.008)	[0.002]	[0.001]	[0.847]	[0.568]
White	7504	0.333	4073	0.318	0.006	0.069	0.088	0.183
	[247]	(0.015)	[225]	(0.020)	[0.004]	[0.041]	[0.078]	[0.204]
Pardo	7504	0.634	4073	0.644	0.152	0.319	0.264	0.657
	[247]	(0.015)	[225]	(0.021)	[0.102]	[0.200]	[0.238]	[0.671]
Black	7504	0.027	4073	0.031	0.004	0.417	0.073	0.016
	[247]	(0.003)	[225]	(0.004)	[0.002]	[0.416]	[0.008]	[0.032]
Receives *Bolsa Famlia*	11285	0.311	6437	0.320	0.305	0.891	0.263	0.239
	[252]	(0.018)	[246]	(0.022)	[0.148]	[0.857]	[0.130]	[0.065]
Receives school transportation	11285	0.126	6437	0.106	0.987	0.664	0.249	0.230
	[252]	(0.015)	[246]	(0.014)	[0.980]	[0.442]	[0.108]	[0.100]

Note: We compare student-level socio-demographics at the beginning of the year of the intervention, i.e., 2016, from that year Rio Grande do Norte school census. Student data regard students enrolled in those grades at the beginning of the school year. Standard errors (SE) are clustered at the school level. Strata (i.e., region and grade) fixed effects are included in all the estimated regressions. We show both standard p-values and p-values computed using randomization inference (RI) with 10,000 repetitions for the whole sample and for each grade.

Table C6: Bounded Treatment Effects on Student Learning, Robust to Missing Data at the School Level

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
All schools					
Naive treatment effect from Table E5	0.007	0.012	-0.007	-0.012	0.023
	(0.047)	(0.053)	(0.057)	(0.045)	(0.043)
Lower Lee bound	-0.001	0.002	-0.017	-0.014	0.014
	(0.041)	(0.047)	(0.048)	(0.040)	(0.041)
Upper Lee bound	0.003	0.007	-0.012	-0.009	0.016
	(0.041)	(0.047)	(0.049)	(0.040)	(0.040)
Imbens and Manski 90% CI	[-0.067, 0.069]	[-0.074, 0.082]	[-0.094, 0.066]	[-0.077, 0.055]	[-0.051, 0.081]
Number of observations	280	280	280	280	280
Number of *selected* observations	270	270	270	270	270
Trimming proportion	0.003	0.003	0.003	0.003	0.003
5th grade – Primary schools					
Naive treatment effect from Table E5	-0.073	-0.070	-0.095	-0.070	-0.068
	(0.088)	(0.098)	(0.093)	(0.088)	(0.086)
Lower Lee bound	-0.093	-0.082	-0.125	-0.069	-0.061
	(0.089)	(0.099)	(0.098)	(0.089)	(0.088)
Upper Lee bound	-0.055	-0.037	-0.079	-0.052	-0.051
	(0.088)	(0.098)	(0.093)	(0.085)	(0.087)
Imbens and Manski 90% CI	[-0.224, 0.074]	[-0.227, 0.107]	[-0.268, 0.057]	[-0.207, 0.081]	[-0.201, 0.087]
Number of observations	97	97	97	97	97
Number of *selected* observations	94	94	94	94	94
Trimming proportion	0.035	0.035	0.035	0.035	0.035
6th grade – Lower secondary schools					
Naive treatment effect from Table E5	0.152**	0.177**	0.170**	0.112**	0.135**
	(0.063)	(0.076)	(0.079)	(0.056)	(0.064)
Lower Lee bound	0.128**	0.177***	0.161**	0.109**	0.131**
	(0.058)	(0.065)	(0.066)	(0.050)	(0.061)
Upper Lee bound	0.152***	0.189***	0.162**	0.114**	0.145**
	(0.054)	(0.065)	(0.068)	(0.051)	(0.060)
Imbens and Manski 90% CI	[0.044, 0.232]	[0.075, 0.291]	[0.052, 0.273]	[0.029, 0.195]	[0.037, 0.238]
Number of observations	105	105	105	105	105
Number of *selected* observations	102	102	102	102	102
Trimming proportion	0.017	0.017	0.017	0.017	0.017
10th grade – Upper secondary schools					
Naive treatment effect from Table E5	-0.019	-0.027	-0.050	-0.046	0.043
	(0.076)	(0.086)	(0.109)	(0.065)	(0.053)
Lower Lee bound	-0.023	-0.033	-0.067	-0.045	0.028
	(0.064)	(0.072)	(0.091)	(0.066)	(0.053)
Upper Lee bound	0.053	0.051	0.038	0.039	0.084
	(0.063)	(0.073)	(0.090)	(0.062)	(0.053)
Imbens and Manski 90% CI	[-0.107, 0.136]	[-0.128, 0.148]	[-0.187, 0.157]	[-0.132, 0.120]	[-0.042, 0.155]
Number of observations	78	78	78	78	78
Number of *selected* observations	75	75	75	75	75
Trimming proportion	0.067	0.067	0.067	0.067	0.067

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school. Point-identified 'naive treatment effects' are estimated by OLS with strata (i.e., region and grade) fixed effects and analytic weights for the number of students enrolled in the grade of interest (see online Appendix Table E5). Partially-identified treatment effect bounds and 90 percent confidence intervals (CI) around such bounds are estimated as described by Lee (2009) and Imbens and Manski (2004), respectively. Robust standard errors for naive treatment effects and bootstrapped – using 10,000 replications – standard errors for bounds in parentheses. The coefficients are expressed in terms of standard deviations from the control group, while mean and standard deviation of the dependent variable refer to the raw values in the control group.

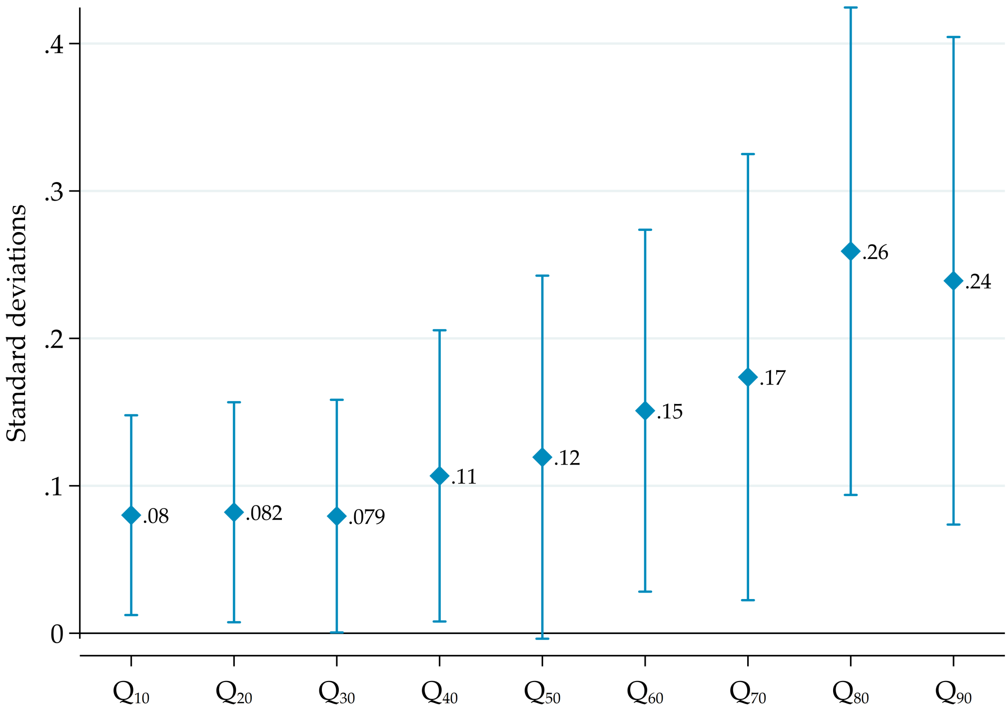
Table C7: Bounded Treatment Effects on Socio-Emotional Skills, Robust to Missing Data at the School Level

	(1) Agreeableness	(2) Conscientiousness	(3) Extraversion	(4) Neuroticism	(5) Openness
All schools					
Naive treatment effect from Table E15	0.075	0.142**	0.103	0.087	0.078
	(0.076)	(0.069)	(0.069)	(0.062)	(0.069)
Lower Lee bound	0.045	0.110	0.036	0.038	0.068
	(0.072)	(0.068)	(0.073)	(0.061)	(0.073)
Upper Lee bound	0.123*	0.176***	0.151**	0.123**	0.132**
	(0.070)	(0.068)	(0.072)	(0.062)	(0.065)
Imbens and Manski 90% CI	[-0.050, 0.216]	[0.018, 0.267]	[-0.058, 0.245]	[-0.042, 0.204]	[-0.031, 0.220]
Number of observations	280	280	280	280	280
Number of *selected* observations	236	236	236	236	236
Trimming proportion	0.118	0.118	0.118	0.118	0.118
5th grade – Primary schools					
Naive treatment effect from Table E15	0.086	0.157	0.050	0.092	-0.010
	(0.125)	(0.120)	(0.118)	(0.102)	(0.121)
Lower Lee bound	0.043	0.152	-0.024	0.027	0.010
	(0.124)	(0.131)	(0.121)	(0.110)	(0.134)
Upper Lee bound	0.115	0.208*	0.091	0.126	0.084
	(0.125)	(0.125)	(0.122)	(0.104)	(0.117)
Imbens and Manski 90% CI	[-0.133, 0.293]	[-0.039, 0.392]	[-0.186, 0.255]	[-0.122, 0.267]	[-0.181, 0.251]
Number of observations	97	97	97	97	97
Number of *selected* observations	85	85	85	85	85
Trimming proportion	0.091	0.091	0.091	0.091	0.091
6th grade – Lower secondary schools					
Naive treatment effect from Table E15	0.098	0.204*	0.274**	0.096	0.182
	(0.120)	(0.111)	(0.124)	(0.096)	(0.124)
Lower Lee bound	0.083	0.161	0.219*	0.057	0.174
	(0.127)	(0.116)	(0.130)	(0.111)	(0.130)
Upper Lee bound	0.173	0.258**	0.323**	0.165	0.235**
	(0.116)	(0.107)	(0.133)	(0.113)	(0.113)
Imbens and Manski 90% CI	[-0.093, 0.334]	[0.003, 0.404]	[0.040, 0.507]	[-0.092, 0.318]	[-0.015, 0.400]
Number of observations	105	105	105	105	105
Number of *selected* observations	88	88	88	88	88
Trimming proportion	0.051	0.051	0.051	0.051	0.051
10th grade – Upper secondary schools					
Naive treatment effect from Table E15	0.039	0.064	0.022	0.071	0.107
	(0.145)	(0.122)	(0.109)	(0.124)	(0.106)
Lower Lee bound	0.015	0.012	-0.051	0.034	0.047
	(0.113)	(0.104)	(0.111)	(0.097)	(0.097)
Upper Lee bound	0.090	0.064	0.074	0.083	0.101
	(0.106)	(0.104)	(0.102)	(0.104)	(0.102)
Imbens and Manski 90% CI	[-0.143, 0.239]	[-0.139, 0.214]	[-0.198, 0.209]	[-0.108, 0.233]	[-0.091, 0.248]
Number of observations	78	78	78	78	78
Number of *selected* observations	63	63	63	63	63
Trimming proportion	0.196	0.196	0.196	0.196	0.196

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school. Point-identified 'naive treatment effects' are estimated by OLS with strata (i.e., region and grade) fixed effects and analytic weights for number of students enrolled in the grade of interest (see online Appendix Table E15). Partially-identified treatment effect bounds and 90 percent confidence intervals (CI) around such bounds are estimated as described by Lee (2009) and Imbens and Manski (2004), respectively. Robust standard errors for naive treatment effects and bootstrapped – using 10,000 replications – standard errors for bounds in parentheses. The coefficients are expressed in terms of standard deviations from the control group, while the mean and standard deviation of the dependent variable refer to the raw values in the control group. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score.

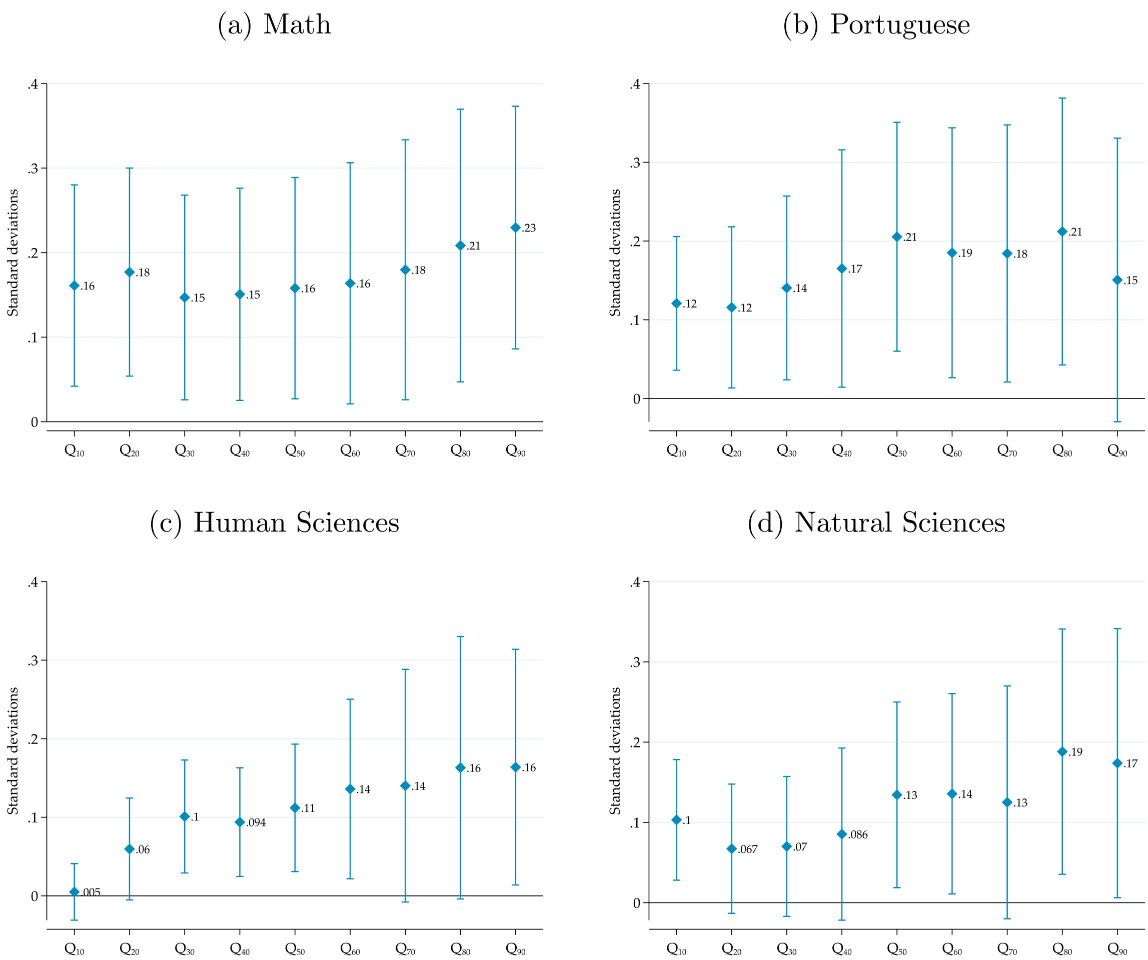
D Auxiliary Results

Figure D1: Quantile Treatment Effect on Average Test Score in 6th Grade



Note: Point estimates of quantile regressions with strata (i.e., region) fixed effects and standard errors clustered at the school level. Confidence intervals are 90 percent. Sample: schools treated at 6th grade. Quantile treatment effects are expressed in terms of standard deviations from the control group.

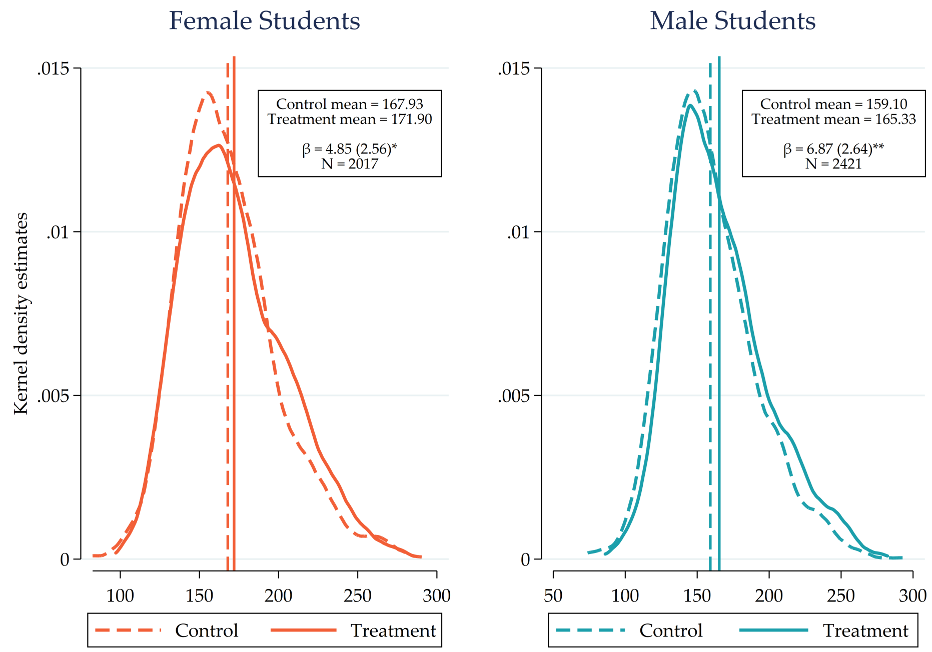
Figure D2: Quantile Treatment Effects in 6th Grade – By Subject



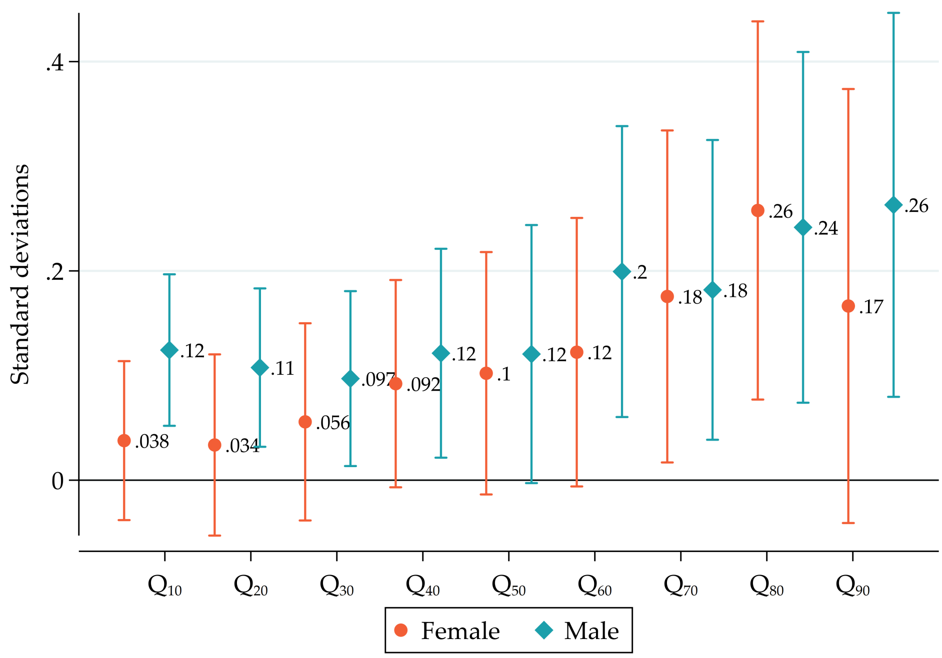
Note: Point estimates of quantile regressions with strata (i.e., region and grade) fixed effects and standard errors clustered at the school level. Confidence intervals are 90 percent. Sample: schools treated at 6th grade. Quantile treatment effects are expressed in terms of standard deviations from the control group.

Figure D3: Impact on Average Test Score by Gender in 6th Grade

(a) Distribution



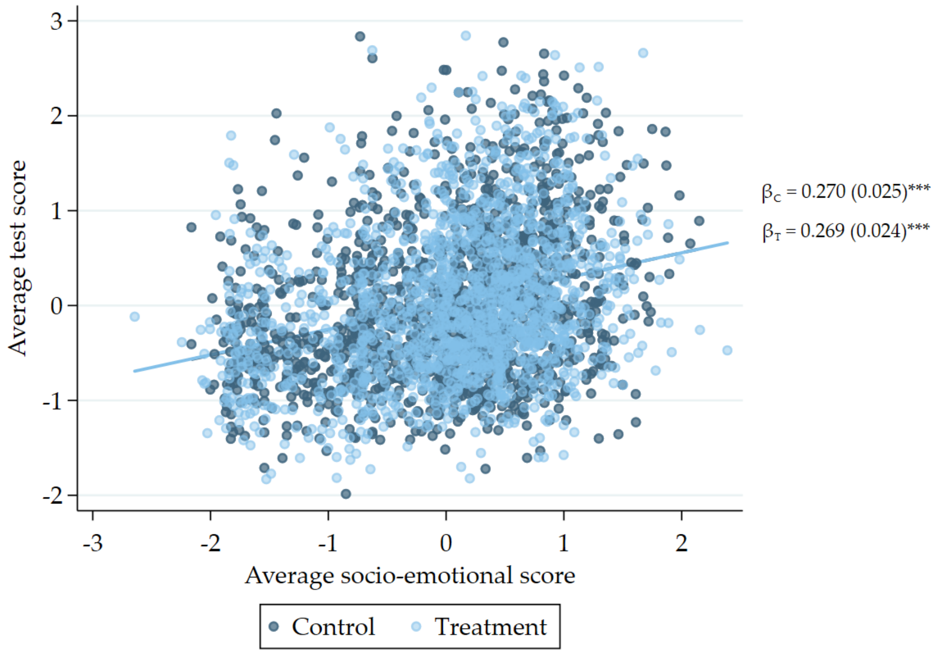
(b) Quantile Treatment Effect



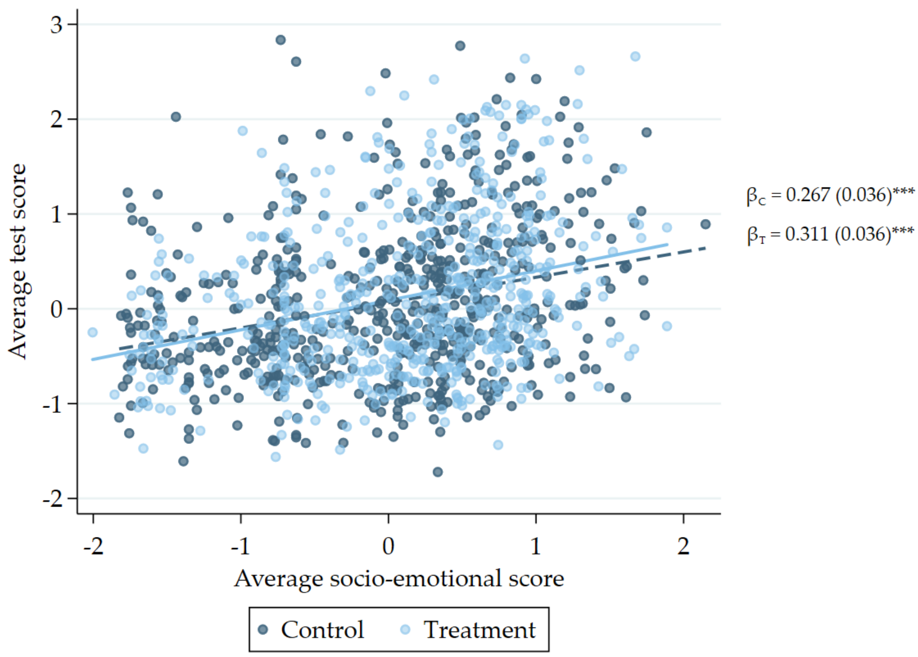
Note: Average test score is the average of standardized test scores in math, Portuguese, human and natural science (range 0-400). Sample: schools treated at 6th grade. Kernel densities are computed using Epanechnikov kernel function. Treatment effects in (a) are estimated through regressions with strata (i.e., region and grade) fixed effects and standard errors clustered at the school level. ** and * indicate significance at the 5 and 10 percent critical level. In (b), we plot point estimates of quantile regressions with 90 percent confidence intervals. Quantile treatment effects are expressed in terms of standard deviations from the control group.

Figure D4: Scatter Plot of Cognitive and Socio-Emotional Skills

(a) All Schools



(b) 6th Grade



Note: Unit of observation: student. The linear fits are estimated for both treatment and control group through an OLS regression with standard errors clustered at the school level. *** indicates significance at the 1 percent critical level. The sample is restricted to students who took the socio-emotional test. 'Average test score' is the average of standardized test scores in math, Portuguese, human and natural science. 'Average socio-emotional score' is the average of standardized scores in agreeableness, conscientiousness, extroversion, neuroticism, and openness. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score. Both variables are expressed in terms of standard deviations from the control group.

Table D1: Effect of Teacher Permanence on Education Outcomes in Brazil

	(1) Age-grade distortion	(2) Passing rate	(3) Repetition rate	(4) Dropout rate
***Ensino Fundamental* – Grades 1-9**				
Teacher permanence index	-1.05***	0.46***	-0.27**	-0.20**
	(0.30)	(0.15)	(0.10)	(0.08)
Number of observations	126739	126223	126223	126223
Number of clusters	27	27	27	27
Adjusted R-squared	0.243	0.157	0.129	0.095
Mean dep. var.	20.50	89.87	7.95	2.18
SD dep. var.	17.37	11.03	8.99	4.89
State fixed effects	✓	✓	✓	✓
***Ensino Medio* – Grades 10-12**				
Teacher permanence index	-4.39***	2.33***	-1.07**	-1.26***
	(0.59)	(0.58)	(0.43)	(0.23)
Number of observations	26505	26552	26552	26552
Number of clusters	27	27	27	27
Adjusted R-squared	0.306	0.153	0.094	0.185
Mean dep. var.	24.64	85.01	9.39	5.61
SD dep. var.	19.18	12.10	8.55	7.53
State fixed effects	✓	✓	✓	✓

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school. Year: 2015. Outcome variables in the column headers. 'Age-grade distortion' is the percentage of students in one grade who are older than the expected age for that grade. 'Teacher permanence index' is the school weighted average of *Indicador de Regularidade Docente*, which takes values between 0 and 5 and is defined as the frequency of a teacher in a school during the last 5 years. The index is standardized so that the coefficients can be interpreted as the effect of one-standard-deviation change in such index. The mean of the 'teacher permanence index' in the sample is 3.04 and the standard deviation is 0.85. All regressions are OLS. Standard errors clustered at the state level in parentheses. The sample is the universe of schools in Brazil. Data are from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP): http://portal.inep.gov.br/indicadores-educacionais.

Table D2: Impact on Student Learning – Standardized Test Scores Rescaled to SAEB

	(1) 5th	(2) 6th	(3) 9th	(4) 10th	(5) 12th
Math					
Treatment	-0.081	0.153**	0.117	-0.002	-0.078
	(0.094)	(0.073)	(0.092)	(0.085)	(0.114)
Number of observations	3065	4226	2118	4744	3257
Number of clusters	96	104	93	78	77
Mean dep. var. control group	164.623	173.851	211.814	220.419	234.287
SD dep. var. control group	46.278	44.655	45.455	37.330	42.500
Portuguese					
Treatment	-0.092	0.133*	0.115	-0.014	-0.029
	(0.087)	(0.076)	(0.085)	(0.107)	(0.112)
Number of observations	3065	4225	2119	4744	3260
Number of clusters	96	104	93	78	77
Mean dep. var. control group	178.878	179.134	218.859	223.413	231.668
SD dep. var. control group	64.768	51.032	49.923	44.412	46.331

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All regressions are OLS with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level. The coefficients are expressed in terms of standard deviations from the control group, while the unconditional mean and standard deviation of the dependent variable refer to the raw values in the control group.

Table D3: Impact on Student Progression Rates – Heterogeneity by Gender

	(1) All	(2) 5th	(3) 6th	(4) 10th
Probability of student passing				
Treatment	0.034	0.016	0.068*	0.023
	(0.023)	(0.025)	(0.035)	(0.042)
Treatment × Male student	0.021	-0.005	-0.002	0.041*
	(0.015)	(0.031)	(0.028)	(0.021)
Male student	-0.109***	-0.055**	-0.109***	-0.128***
	(0.011)	(0.023)	(0.019)	(0.014)
Constant	0.658***	0.821***	0.645***	0.593***
	(0.017)	(0.019)	(0.022)	(0.030)
Total effect on male students: Treatment + Treatment × male student				
$\sum \hat{\beta}$	0.055	0.011	0.067	0.065
P-value	0.019	0.757	0.047	0.136
Probability of student dropping out				
Treatment	-0.001	0.002	-0.049**	0.026
	(0.015)	(0.015)	(0.021)	(0.027)
Treatment × Male student	-0.015	0.000	0.012	-0.031
	(0.012)	(0.018)	(0.020)	(0.020)
Male student	0.050***	0.022*	0.029**	0.072***
	(0.008)	(0.012)	(0.014)	(0.014)
Constant	0.141***	0.072***	0.120***	0.189***
	(0.010)	(0.011)	(0.017)	(0.017)
Total effect on male students: Treatment + Treatment × male student				
$\sum \hat{\beta}$	-0.016	0.002	-0.037	-0.005
P-value	0.323	0.891	0.071	0.866
Probability of student repeating				
Treatment	-0.033*	-0.018	-0.019	-0.049
	(0.018)	(0.019)	(0.033)	(0.030)
Treatment × Male student	-0.006	0.005	-0.011	-0.010
	(0.014)	(0.024)	(0.025)	(0.022)
Male student	0.059***	0.033*	0.080***	0.057***
	(0.010)	(0.019)	(0.018)	(0.015)
Constant	0.201***	0.108***	0.235***	0.217***
	(0.014)	(0.015)	(0.018)	(0.024)
Total effect on male students: Treatment + Treatment × male student				
$\sum \hat{\beta}$	-0.039	-0.014	-0.030	-0.060
P-value	0.033	0.581	0.306	0.076

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data constructed from Rio Grande do Norte 2016 and 2017 censuses. Unit of observation: student. Outcome variables in the panel headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level in parentheses.

Table D4: Impact on Student Progression Rates – Heterogeneity by Passing Rate at Baseline

	Grade level				*Student level*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	5th	6th	10th	All	5th	6th	10th
	Passing							
Treatment	4.30^{*}	3.26	5.42	3.72	5.29^{**}	5.22	5.25	5.10
	(2.54)	(4.02)	(4.65)	(4.67)	(2.64)	(4.38)	(3.86)	(4.38)
Treatment × High passing rate at baseline	-0.44	-1.60	1.61	-1.34	-3.11	-7.04	0.87	-3.22
	(3.25)	(4.77)	(6.01)	(6.16)	(3.84)	(5.14)	(5.40)	(7.38)
High passing rate at baseline	14.79^{***}	11.11^{***}	15.33^{***}	18.24^{***}	16.87^{***}	13.00^{***}	14.18^{***}	20.32^{***}
	(2.18)	(3.43)	(3.76)	(4.30)	(2.55)	(3.61)	(3.38)	(4.76)
Constant	63.84^{***}	77.49^{***}	56.81^{***}	56.82^{***}	53.26^{***}	72.05^{***}	52.56^{***}	45.85^{***}
	(1.63)	(2.79)	(2.57)	(3.21)	(1.64)	(2.93)	(2.07)	(2.69)
Total effect on schools with high passing rate at baseline: Treatment + Treatment × high-passing dummy								
$\sum \hat{\beta}$	3.863	1.653	7.033	2.383	2.187	-1.817	6.113	1.881
P-value	0.057	0.517	0.071	0.547	0.438	0.492	0.111	0.758
	Dropout							
Treatment	-0.98	0.94	-2.49	-1.14	0.40	-1.37	-4.74^{*}	4.15
	(1.28)	(1.22)	(1.97)	(3.32)	(2.07)	(2.17)	(2.79)	(3.59)
Treatment × High passing rate at baseline	1.65	-2.10	2.44	5.20	-2.36	2.93	1.55	-7.46
	(1.61)	(1.55)	(2.32)	(4.40)	(2.52)	(2.59)	(3.23)	(4.69)
High passing rate at baseline	-4.08^{***}	-0.45	-5.23^{***}	-6.75^{**}	-5.69^{***}	-5.59^{***}	-6.18^{**}	-5.43
	(1.13)	(1.13)	(1.67)	(3.03)	(2.05)	(1.89)	(2.67)	(3.94)
Constant	8.14^{***}	2.32^{***}	9.16^{***}	13.59^{***}	19.04^{***}	11.36^{***}	16.22^{***}	24.43^{***}
	(0.95)	(0.79)	(1.51)	(2.52)	(1.58)	(1.61)	(2.26)	(2.71)
Total effect on schools with high passing rate at baseline: Treatment + Treatment × high-passing dummy								
$\sum \hat{\beta}$	0.672	-1.153	-0.057	4.056	-1.951	1.556	-3.192	-3.303
P-value	0.495	0.227	0.964	0.165	0.176	0.282	0.063	0.292
	Repetition							
Treatment	-3.32	-4.20	-2.93	-2.57	-5.71^{***}	-3.85	-0.53	-9.26^{***}
	(2.44)	(3.78)	(4.18)	(4.94)	(2.11)	(3.17)	(3.76)	(3.07)
Treatment × High passing rate at baseline	-1.21	3.70	-4.05	-3.86	5.49^{*}	4.12	-2.37	10.68^{**}
	(3.17)	(4.50)	(5.50)	(6.72)	(3.07)	(3.78)	(5.30)	(5.34)
High passing rate at baseline	-10.71^{***}	-10.67^{***}	-10.10^{***}	-11.49^{***}	-11.18^{***}	-7.37^{***}	-8.02^{***}	-14.89^{***}
	(2.08)	(3.22)	(3.57)	(4.21)	(1.96)	(2.58)	(3.00)	(3.32)
Constant	28.02^{***}	20.19^{***}	34.03^{***}	29.59^{***}	27.68^{***}	16.55^{***}	31.22^{***}	29.71^{***}
	(1.50)	(2.65)	(2.50)	(2.66)	(1.39)	(2.11)	(2.02)	(2.14)
Total effect on schools with high passing rate at baseline: Treatment + Treatment × high-passing dummy								
$\sum \hat{\beta}$	-4.536	-0.500	-6.976	-6.439	-0.229	0.269	-2.902	1.421
P-value	0.024	0.838	0.053	0.149	0.918	0.894	0.427	0.749

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. School-level data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses, respectively. Outcome variables in the panel headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors for school-level regressions and standard errors clustered at the school level for student-level regressions in parentheses. The coefficients are expressed in terms of percentage points and the mean and standard deviation of the dependent variable in the control group are unconditional.

Table D5: Impact of 6th Grade Repetition on Student Achievement

	Dropout			Years of completed schooling		
	(1)	(2)	(3)	(4)	(5)	(6)
Repetition in 2011	0.280***	0.292***	0.214***	-2.021***	-2.072***	-1.679***
	(0.007)	(0.005)	(0.005)	(0.035)	(0.022)	(0.023)
Number of observations	73010	73007	72994	73010	73007	72994
Number of clusters	1154	1151	2680	1154	1151	2680
Adjusted R-squared	0.065	0.002	0.210	0.149	0.081	0.325
School fixed effects		✓	✓		✓	✓
Class fixed effects			✓			✓

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are constructed from 2011-2016 Rio Grande do Norte censuses. Outcome variables in the column headers. All regressions are OLS. Standard errors clustered at the school – columns (1), (2), (4), (5) – or class – columns (3), (6) – level in parentheses. The sample is the universe of 6th grade students at public schools in Rio Grande do Norte. 'Dropout' is a dummy variable equal to 1 if the student dropped out in one year between 2011 and 2016, and 0 otherwise. 'Years of completed schooling' is taken in the last year in which the student is in the census database. When the student drops out, we consider his/her last grade as its level of completed schooling.

Table D6: Impact of Teacher Participation on Teacher Retention

	(1) All	(2) 5th	(3) 6th	(4) 10th
1. First stage				
Treatment	0.372***	0.547***	0.343***	0.371***
	(0.026)	(0.057)	(0.037)	(0.041)
2. Second Stage				
Teacher participation	0.097	-0.193	0.186*	0.094
	(0.080)	(0.141)	(0.108)	(0.133)
Number of observations	1848	162	784	902
Number of clusters	256	76	104	76
Effective F-statistic	204.862	92.251	85.894	82.836
Partial R-squared	0.248	0.449	0.231	0.235

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from Rio Grande do Norte 2016 and 2017 teacher censuses and State Secretary of Education and Culture (SEEC). Unit of observation: teacher. Treatment-on-the-treated (TOT) coefficients are from two-stage least squares regressions with strata (i.e., region and grade) fixed effects, where the school treatment assignment is used as included instrument for teacher participation to project activities. Standard errors clustered at the school level in parentheses. Effective F-statistics are based on the weak instrument test of Montiel Olea and Pflueger (2013) that is robust to clustering.

Table D7: Drivers of Implementation

	(1)	(2)	(3)	(4)
Number of enrolled students in project grades	-0.000	-0.000	-0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Quality score of expression of interest	0.002	0.003	0.002	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Project grant per student	-0.000	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
School infrastucture index	-0.078*	-0.089**	-0.085*	-0.048
	(0.042)	(0.044)	(0.043)	(0.033)
Urban	-0.080	-0.074	-0.070	0.007
	(0.093)	(0.090)	(0.091)	(0.070)
Distance to Natal (km)	0.002***	0.003**	0.003***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Low teacher retention at baseline		-0.024	-0.019	0.004
		(0.071)	(0.069)	(0.055)
Passing rate at baseline			-0.001	-0.001
			(0.002)	(0.002)
Dropout rate at baseline			0.001	0.000
			(0.006)	(0.003)
School has clearance certificate				0.499***
				(0.092)
Number of observations	123	108	108	108
Adjusted R-squared	0.188	0.189	0.174	0.506
Mean dep. var.	0.826	0.824	0.824	0.824
SD dep. var.	0.340	0.350	0.350	0.350

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from 2015 Rio Grande do Norte school census, State Secretary of Education and Culture (SEEC), and *Sistema Integrado de Gesto da Educao* (SIGEduc). 'Implementation' is defined as the ratio of the number of activities that were implemented over the number of planned activities described in the work plan. 'School infrastructure index' is constructed through the principal component analysis of the following dummy variables: whether the school has internet, a library, a science lab, and is located in an urban area. Unit of observation: school. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors are in parentheses.

Table D8: Correlation between Student Outcomes and Quality of Proposal

	Passing				*Average test score*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
Expression of interest score	0.004	0.007	0.005	-0.001	0.001	0.001	0.000	0.002
	(0.002)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)	(0.002)	(0.003)
Number of observations	5914	1590	2177	2147	7913	1801	2539	3573
Mean dep. var.	0.011	-0.099	0.103	-0.002	0.653	0.802	0.649	0.581
SD dep. var.	0.793	0.818	0.820	0.734	0.476	0.398	0.477	0.494

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from Rio Grande do Norte census and State Secretary of Education and Culture (SEEC). Unit of observation: student. Outcome variables in the panel headers. 'Passing' is a dummy variable (0/1), while 'average test score' is the average of standardized test scores in math, Portuguese, human and natural science and is expressed in terms of standard deviations from the control group. 'Expression of interest' is a score between 0 and 100. All regressions are OLS with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level are in parentheses. The mean and standard deviation of the dependent variable are unconditional.

Table D9: Impact on Probability of School Obtaining the Clearance Certificate

	(1) All	(2) 5th	(3) 6th	(4) 10th
Treatment	0.410***	0.393***	0.421***	0.414***
	(0.051)	(0.088)	(0.083)	(0.095)
Number of observations	278	96	104	78
Mean dep. var. control group	0.364	0.346	0.424	0.302
SD dep. var. control group	0.483	0.480	0.498	0.465
Treatment effect comparisons by grade:				
			$\widehat{\beta_{6th}} - \widehat{\beta_{5th}} = 0.028$ T-test p-value = 0.819	
				$\widehat{\beta_{6th}} - \widehat{\beta_{10th}} = 0.007$ T-test p-value = 0.956

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from the State Secretary of Education and Culture (SEEC). Unit of observation: school. 'Treatment effect comparisons by grade' are based on the regression in column (1) with grade interaction terms. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors in parentheses.

Table D10: Impact on Other Grades in 6th Grade Treated Schools

	(1) 6th	(2) 7th	(3) 8th	(4) 9th
Panel A – Teacher level				
	Probability of teacher staying in the same school			
Treatment	0.064*	0.078**	0.044	0.049
	(0.037)	(0.037)	(0.037)	(0.039)
Number of observations	784	792	759	682
Number of clusters	104	103	99	93
Mean dep. var. control group	0.691	0.697	0.691	0.688
SD dep. var. control group	0.463	0.460	0.463	0.464
Panel B – Student level				
	Probability of student being promoted			
Treatment	0.070**	0.014	-0.005	0.010
	(0.031)	(0.032)	(0.029)	(0.036)
Number of observations	5490	4465	3294	2883
Number of clusters	104	103	99	93
Mean dep. var. control group	0.587	0.669	0.809	0.778
SD dep. var. control group	0.492	0.471	0.393	0.416
	Probability of student dropping out			
Treatment	-0.043**	-0.029	-0.013	-0.038*
	(0.018)	(0.018)	(0.018)	(0.022)
Number of observations	5494	4473	3303	2889
Number of clusters	104	103	99	93
Mean dep. var. control group	0.135	0.136	0.114	0.151
SD dep. var. control group	0.342	0.343	0.317	0.358
	Probability of student repeating			
Treatment	-0.027	0.015	0.018	0.028
	(0.028)	(0.027)	(0.017)	(0.021)
Number of observations	5490	4465	3294	2883
Number of clusters	104	103	99	93
Mean dep. var. control group	0.277	0.195	0.077	0.071
SD dep. var. control group	0.448	0.396	0.266	0.257

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from Rio Grande do Norte 2016 and 2017 teacher and student censuses. Unit of observation: teacher in the first panel and student in the other panels. Sample: schools treated at 6th grade. All outcome variables (in the panel headers) are dummy variables and regressions are linear probability model with strata (i.e., region) fixed effects. Standard errors clustered at the school level in parentheses.

Table D11: Impact on Student Progression Rates in 6th Grade Treated Schools – Spillover to Other Grades

	Grade level				*Student level*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	6th	7th	8th	9th	6th	7th	8th	9th
				Passing				
Treatment	8.46**	-0.16	-1.60	-4.62	7.00**	1.41	-0.46	0.95
	(3.30)	(2.93)	(2.64)	(3.23)	(3.10)	(3.18)	(2.88)	(3.61)
Number of observations	104	103	99	93	5490	4465	3294	2883
Number of clusters					104	103	99	93
Mean dep. var. control group	63.56	72.63	86.08	86.24	58.73	66.93	80.92	77.79
SD dep. var. control group	17.05	13.46	12.32	10.94	49.24	47.06	39.30	41.58
				Dropout				
Treatment	-1.61	-0.31	0.05	4.15	-4.35**	-2.88	-1.32	-3.78*
	(1.27)	(1.38)	(1.42)	(2.76)	(1.82)	(1.78)	(1.75)	(2.25)
Number of observations	104	103	99	93	5494	4473	3303	2889
Number of clusters					104	103	99	93
Mean dep. var. control group	6.84	5.67	4.73	5.69	13.55	13.57	11.37	15.10
SD dep. var. control group	7.15	6.57	7.16	7.23	34.23	34.25	31.75	35.82
				Repetition				
Treatment	-6.85**	0.48	1.55	0.47	-2.65	1.46	1.84	2.84
	(2.91)	(2.89)	(2.13)	(2.08)	(2.81)	(2.74)	(1.74)	(2.08)
Number of observations	104	103	99	93	5490	4465	3294	2883
Number of clusters					104	103	99	93
Mean dep. var. control group	29.59	21.70	9.19	8.07	27.72	19.48	7.65	7.08
SD dep. var. control group	14.91	12.37	10.02	9.51	44.77	39.61	26.60	25.66

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. School-level data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses, respectively. Sample: schools treated at 6th grade. All regressions are OLS with strata (i.e., region) fixed effects. Robust standard errors for school-level regressions and standard errors clustered at the school level for student-level regressions in parentheses.

Table D12: Impact on Student Progression Rates in 6th Grade Treated Schools by Teacher Retention at Baseline – Spillover to Other Grades

	(1) 7th	(2) 8th	(3) 9th
Probability of student passing			
Treatment	0.052	-0.046	0.006
	(0.046)	(0.030)	(0.043)
Treatment × Low teacher retention at baseline	-0.055	0.070	0.014
	(0.061)	(0.054)	(0.069)
Low teacher retention at baseline	-0.051	-0.092**	-0.045
	(0.041)	(0.038)	(0.034)
Constant	0.695***	0.862***	0.791***
	(0.029)	(0.019)	(0.019)
Number of observations	4299	3205	2828
Number of clusters	97	95	90
Total effect: Treatment + Treatment × Low teacher retention at baseline			
$\sum \hat{\beta}$	-0.003	0.024	0.020
P-value	0.948	0.603	0.721
Probability of student dropping out			
Treatment	-0.056***	-0.014	-0.034
	(0.018)	(0.019)	(0.031)
Treatment × Low teacher retention at baseline	0.030	0.000	-0.007
	(0.030)	(0.033)	(0.044)
Low teacher retention at baseline	0.024	0.039	-0.003
	(0.024)	(0.027)	(0.025)
Constant	0.126***	0.093***	0.159***
	(0.017)	(0.015)	(0.018)
Number of observations	4307	3214	2833
Number of clusters	97	95	90
Total effect: Treatment + Treatment × Low teacher retention at baseline			
$\sum \hat{\beta}$	-0.027	-0.014	-0.041
P-value	0.289	0.615	0.205
Probability of student being retained			
Treatment	0.004	0.062***	0.027
	(0.045)	(0.022)	(0.021)
Treatment × Low teacher retention at baseline	0.026	-0.071**	-0.006
	(0.055)	(0.033)	(0.038)
Low teacher retention at baseline	0.026	0.054**	0.047**
	(0.034)	(0.023)	(0.021)
Constant	0.179***	0.044***	0.050***
	(0.027)	(0.015)	(0.010)
Number of observations	4299	3205	2828
Number of clusters	97	95	90
Total effect: Treatment + Treatment × Low teacher retention at baseline			
$\sum \hat{\beta}$	0.030	-0.010	0.022
P-value	0.372	0.693	0.504

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Student and teacher data are from Rio Grande do Norte censuses. Unit of observation: student. Sample: schools treated at 6th grade. $\sum \hat{\beta}$ is the sum of the treatment effect with the interaction variable coefficient. The p-value refers to the null hypothesis $\sum \hat{\beta} = 0$. All regressions are OLS with strata (i.e., region) fixed effects. The coefficients on progression are expressed in terms of percentage points. Standard errors clustered at the school level in parentheses.

Table D13: Impact on Student Learning and Progression by Teacher Turnover at Baseline

	Learning	*Progression*		
	(1) Average test score	(2) Passed	(3) Dropped out	(4) Retained
6th grade – Lower seconday schools				
Treatment	0.016	0.079*	-0.025	-0.054
	(0.080)	(0.043)	(0.027)	(0.037)
Treatment × Low teacher retention at baseline	0.261**	-0.023	-0.043	0.066
	(0.119)	(0.058)	(0.037)	(0.053)
Low teacher retention at baseline	-0.200***	-0.046	0.045	0.000
	(0.063)	(0.040)	(0.032)	(0.034)
Constant	0.062	0.613***	0.115***	0.272***
	(0.039)	(0.029)	(0.024)	(0.023)
Number of observations	4333	5261	5265	5261
Number of clusters	94	98	98	98
Total effect: Treatment + Treatment × Low teacher retention at baseline				
$\sum \hat{\beta}$	0.276	0.055	-0.067	0.012
P-value	0.002	0.186	0.008	0.752
10th grade – Upper secondary schools				
Treatment	-0.118	0.079	0.027	-0.106***
	(0.112)	(0.052)	(0.037)	(0.038)
Treatment × Low teacher retention at baseline	0.196	-0.065	-0.029	0.094*
	(0.122)	(0.077)	(0.054)	(0.052)
Low teacher retention at baseline	-0.267**	0.097*	0.005	-0.103***
	(0.109)	(0.050)	(0.040)	(0.038)
Constant	0.137	0.479***	0.220***	0.301***
	(0.105)	(0.033)	(0.029)	(0.031)
Number of observations	5070	8157	8159	8157
Number of clusters	73	78	78	78
Total effect: Treatment + Treatment × Low teacher retention at baseline				
$\sum \hat{\beta}$	0.078	0.014	-0.003	-0.012
P-value	0.217	0.805	0.945	0.753

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. Outcome variables in the column headers. 'Average test score' is the average of standardized test scores in math, Portuguese, human and natural science. Student-level data on progression are from Rio Grande do Norte census. Teacher data are from Rio Grande do Norte 2016 and 2017 teacher censuses. Unit of observation: student. $\sum \hat{\beta}$ is the sum of the treatment effect with the interaction variable coefficient. The p-value refers to the null hypothesis $\sum \hat{\beta} = 0$. All regressions are OLS with strata (i.e., region and grade) fixed effects. The coefficients on learning is expressed in terms of standard deviations from the control group, while the coefficients on progression are expressed in terms of percentage points. Standard errors clustered at the school level in parentheses.

Table D14: Correlation between Outcomes and Project Grant per Student

	Passing				*Average test score*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
Grant per student (BRL)	-0.00001	-0.00003	0.00004	-0.00004	-0.00009	-0.00043	-0.00010	-0.00002
	(0.00003)	(0.00009)	(0.00003)	(0.00004)	(0.00005)	(0.00031)	(0.00008)	(0.00007)
Number of observations	7913	1801	2539	3573	5914	1590	2177	2147
Mean dep. var.	0.653	0.802	0.649	0.581	0.011	-0.099	0.103	-0.002
SD dep. var.	0.476	0.398	0.477	0.494	0.793	0.818	0.820	0.734

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from Rio Grande do Norte census and State Secretary of Education and Culture (SEEC). Unit of observation: student. Outcome variables in the panel headers. 'Passing' is a dummy variable (0/1), while 'average test score' is the average of standardized test scores in math, Portuguese, human, and natural science and is expressed in terms of standard deviations from the control group. 'Grant per student' is in Brazilian *reais* (BRL). All regressions are OLS with strata (i.e., region and grade) fixed effects. Standard errors clustered at the school level are in parentheses. The mean and standard deviation of the dependent variable are unconditional.

Table D15: Impact on Share of Federal Funds Disbursed

	(1) All	(2) 5th	(3) 6th	(4) 10th
Intervention school year (2016)				
Treatment	4.25	-1.26	13.23	-0.92
	(14.91)	(26.57)	(29.70)	(14.19)
Number of observations	277	95	104	78
Mean dep. var. control group	4.05	1.34	-13.88	31.85
SD dep. var. control group	155.38	166.59	188.12	64.97
Post-intervention school year (2017)				
Treatment	23.87	5.09	60.45**	-1.75
	(16.73)	(37.02)	(26.57)	(14.78)
Number of observations	277	95	104	78
Mean dep. var. control group	11.77	4.71	-3.96	41.73
SD dep. var. control group	180.74	224.61	195.52	63.58

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from *Fundo Nacional de Desenvolvimento da Educao* (FNDE). Unit of observation: school. The outcome variable is the percentage of federal funds from the *Programa Dinheiro Direto na Escola* (PDDE) program disbursed to the schools. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors in parentheses.

E Robustness

Table E1: Impact on Student Learning – Controlling for Students' Characteristics

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
All schools					
Treatment	0.025	0.048	0.011	-0.010	0.047
	(0.045)	(0.053)	(0.055)	(0.042)	(0.042)
Number of observations	7501	6735	6738	6473	6468
Number of clusters	247	246	246	246	246
Mean dep. var. control group	183.110	172.557	191.239	183.849	183.181
SD dep. var. control group	41.531	47.868	53.273	49.299	42.794
5th grade – Primary schools					
Treatment	-0.083	-0.082	-0.134*	-0.076	-0.074
	(0.073)	(0.087)	(0.073)	(0.078)	(0.078)
Number of observations	1865	1719	1721	1748	1746
Number of clusters	89	89	89	89	89
Mean dep. var. control group	161.031	161.372	178.781	156.159	151.470
SD dep. var. control group	35.929	44.061	58.953	36.943	28.855
6th grade – Lower secondary schools					
Treatment	0.136**	0.180**	0.152*	0.084	0.118*
	(0.061)	(0.069)	(0.080)	(0.059)	(0.063)
Number of observations	3002	2679	2680	2749	2748
Number of clusters	97	96	96	97	97
Mean dep. var. control group	163.481	152.810	172.690	161.140	171.146
SD dep. var. control group	32.253	43.631	47.617	36.038	35.516
10th grade – Upper secondary schools					
Treatment	-0.042	-0.033	-0.063	-0.091	0.020
	(0.076)	(0.093)	(0.100)	(0.067)	(0.058)
Number of observations	2634	2337	2337	1976	1974
Number of clusters	61	61	61	60	60
Mean dep. var. control group	218.601	201.511	219.796	236.099	225.179
SD dep. var. control group	27.325	40.126	41.401	27.222	24.689

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All regressions are OLS with strata fixed effects and control for students' characteristics, such as age, gender and race dummies (white, indigenous, black, or *pardo*), whether they receive *Bolsa Família*, and whether they use school transportation. Standard errors clustered at the school level in parentheses.

Table E2: Impact on Student Learning – Blocked Difference-in-Means

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
All schools					
Treatment	-0.010	-0.018	-0.022	-0.010	0.024
	(0.058)	(0.065)	(0.078)	(0.046)	(0.041)
Number of observations	12760	11366	11365	10885	10879
Number of clusters	264	264	264	264	264
Mean dep. var. control group	184.052	172.693	190.234	186.477	185.329
SD dep. var. control group	41.081	46.528	52.637	49.517	42.864
5th grade – Primary schools					
Treatment	-0.123	-0.124	-0.134	-0.121	-0.130
	(0.097)	(0.108)	(0.100)	(0.098)	(0.090)
Number of observations	3179	2885	2885	2977	2978
Number of clusters	92	92	92	92	92
Mean dep. var. control group	157.452	157.540	173.368	154.288	149.499
SD dep. var. control group	36.022	43.798	60.456	37.359	28.700
6th grade – Lower secondary schools					
Treatment	0.140**	0.154*	0.163*	0.095	0.124*
	(0.069)	(0.081)	(0.083)	(0.061)	(0.066)
Number of observations	4511	4014	4013	4134	4131
Number of clusters	99	99	99	99	99
Mean dep. var. control group	162.845	151.930	172.451	160.075	170.685
SD dep. var. control group	31.523	42.024	47.502	35.775	35.164
10th grade – Upper secondary schools					
Treatment	-0.061	-0.084	-0.094	-0.043	0.022
	(0.092)	(0.102)	(0.133)	(0.075)	(0.059)
Number of observations	5070	4467	4467	3774	3770
Number of clusters	73	73	73	73	73
Mean dep. var. control group	215.446	198.009	214.086	233.701	223.680
SD dep. var. control group	26.923	38.838	41.371	26.369	23.650

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. Coefficients are sample-weighted average treatment effects of the within-block difference-in-means (Blocked DIM). Standard errors clustered at the school level are in parentheses.

Table E3: Impact on Student Learning – Interaction-Weighted Estimator (IWE)

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
5th grade – Primary schools					
Treatment	-0.064	-0.064	-0.088	-0.067	-0.071
	(0.084)	(0.094)	(0.088)	(0.084)	(0.081)
Percentage difference between IWE and OLS	-4.617	-4.112	-2.815	-4.536	-4.150
P-value for joint test of equality between IWE and OLS	0.107	0.064	0.146	0.103	0.038
P-value for joint Wald Test for interactions	0.556	0.557	0.655	0.547	0.401
6th grade – Lower secondary schools					
Treatment	0.145**	0.177**	0.157**	0.102*	0.121**
	(0.061)	(0.072)	(0.073)	(0.054)	(0.062)
Percentage difference between IWE and OLS	-0.679	-0.101	-0.788	-0.577	-1.095
P-value for joint test of equality between IWE and OLS	0.703	0.723	0.459	0.756	0.661
P-value for joint Wald Test for interactions	0.956	0.931	0.872	0.957	0.924

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All estimates are IWE as in Gibbons et al. (2018). Standard errors clustered at the school level are in parentheses. We only show results for 5th and 6th grades, because 10th grade has one stratum with no variation in treatment assignment.

Table E4: Impact on Student Learning – Regression-Weighted Estimator (RWE)

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
All schools					
Treatment	0.034	0.045	0.031	0.011	0.044
	(0.044)	(0.050)	(0.056)	(0.038)	(0.038)
Percentage difference between RWE and OLS	5.428	8.406	9.503	-11.609	0.050
P-value for joint test of equality between RWE and OLS	0.423	0.134	0.334	0.394	0.991
5th grade – Primary schools					
Treatment	-0.065	-0.065	-0.088	-0.067	-0.071
	(0.087)	(0.097)	(0.090)	(0.087)	(0.084)
Percentage difference between RWE and OLS	-4.334	-3.803	-2.605	-4.230	-3.891
P-value for joint test of equality between RWE and OLS	0.091	0.131	0.175	0.109	0.056
6th grade – Lower secondary schools					
Treatment	0.145**	0.177**	0.157**	0.102*	0.121**
	(0.061)	(0.073)	(0.075)	(0.054)	(0.062)
Percentage difference between RWE and OLS	-0.663	-0.107	-0.752	-0.569	-1.053
P-value for joint test of equality between RWE and OLS	0.345	0.790	0.165	0.518	0.262
10th grade – Upper secondary schools					
Treatment	-0.004	-0.004	-0.006	-0.029	0.051
	(0.076)	(0.086)	(0.109)	(0.061)	(0.053)
Percentage difference between RWE and OLS	-59.928	-74.784	-62.107	10.203	0.484
P-value for joint test of equality between RWE and OLS	0.256	0.063	0.195	0.557	0.962

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All estimates are RWE as in Gibbons et al. (2018). Standard errors clustered at the school level in parentheses.

Table E5: Impact on Student Learning – School Level Regressions

	(1) Average	(2) Math	(3) Portuguese	(4) Human Sciences	(5) Natural Sciences
All schools					
Treatment	0.007	0.012	-0.007	-0.012	0.023
	(0.047)	(0.053)	(0.056)	(0.045)	(0.043)
Number of observations	263	263	263	263	263
Mean dep. var. control group	174.729	165.605	182.673	177.650	177.633
SD dep. var. control group	28.554	25.473	27.435	36.662	32.436
5th grade – Primary schools					
Treatment	-0.073	-0.070	-0.095	-0.070	-0.068
	(0.088)	(0.098)	(0.093)	(0.088)	(0.086)
Number of observations	92	92	92	92	92
Mean dep. var. control group	155.323	154.385	169.916	152.488	147.769
SD dep. var. control group	17.048	19.485	25.530	15.121	12.779
6th grade – Lower secondary schools					
Treatment	0.152**	0.177**	0.170**	0.112**	0.135**
	(0.063)	(0.076)	(0.079)	(0.056)	(0.064)
Number of observations	99	99	99	99	99
Mean dep. var. control group	162.781	152.310	171.844	159.607	170.514
SD dep. var. control group	12.395	14.902	17.670	9.929	12.829
10th grade – Upper secondary schools					
Treatment	-0.019	-0.027	-0.050	-0.046	0.043
	(0.076)	(0.086)	(0.108)	(0.065)	(0.053)
Number of observations	72	72	72	72	72
Mean dep. var. control group	215.228	197.963	213.461	233.734	224.181
SD dep. var. control group	8.703	11.379	13.835	7.299	5.995

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school. All regressions are OLS with strata (i.e., region and grade) fixed effects and analytic weights for the number of students enrolled in the grade of interest. Robust standard errors are in parentheses. The coefficients are expressed in terms of standard deviations from the control group, while the mean and standard deviation of the dependent variable refer to the raw values in the control group.

Table E6: Impact on Student Progression – Controlling for Students' Characteristics

	(1) All	(2) 5th	(3) 6th	(4) 10th
	Passing			
Treatment	0.041**	0.012	0.040	0.044
	(0.020)	(0.022)	(0.026)	(0.037)
Number of observations	17276	3629	5490	8157
Number of clusters	277	95	104	78
Mean dep. var. control group	0.599	0.796	0.587	0.528
SD dep. var. control group	0.490	0.403	0.492	0.499
	Dropout			
Treatment	-0.005	0.004	-0.023*	0.012
	(0.012)	(0.012)	(0.014)	(0.023)
Number of observations	17290	3637	5494	8159
Number of clusters	277	95	104	78
Mean dep. var. control group	0.168	0.082	0.135	0.224
SD dep. var. control group	0.374	0.274	0.342	0.417
	Repetition			
Treatment	-0.036**	-0.016	-0.017	-0.056*
	(0.017)	(0.017)	(0.028)	(0.029)
Number of observations	17276	3629	5490	8157
Number of clusters	277	95	104	78
Mean dep. var. control group	0.233	0.122	0.277	0.248
SD dep. var. control group	0.422	0.327	0.448	0.432

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. Student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses. Outcome variables in the panel headers. All regressions are OLS with strata fixed effects and control for students' characteristics, such as age, gender, and race dummies (white, indigenous, black, or *pardo*), and whether they use school public transportation. The mean and standard deviation of the dependent variable in the control group is unconditional. Standard errors clustered at the school level are in parentheses.

Table E7: Impact on Progression – Blocked Difference-in-Means

	Grade level				*Student level*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
					Passing			
Treatment	5.37***	4.04	9.18***	1.77	4.80	2.54	6.68**	4.59
	(1.88)	(2.69)	(3.30)	(4.14)	(3.13)	(3.16)	(3.30)	(4.66)
Number of observations	277	95	104	78	17276	3629	5490	8157
Number of clusters					277	95	104	78
Mean dep. var. control group	70.97	83.55	63.56	66.22	59.91	79.60	58.73	52.81
SD dep. var. control group	18.04	13.64	17.05	16.23	49.01	40.31	49.24	49.93
					Dropout			
Treatment	-0.50	-0.35	-2.43*	2.19	-1.01	-0.76	-5.59***	0.57
	(0.82)	(0.83)	(1.27)	(2.43)	(2.15)	(1.64)	(2.11)	(3.22)
Number of observations	277	95	104	78	17290	3637	5494	8159
Number of clusters					277	95	104	78
Mean dep. var. control group	6.19	2.09	6.84	10.17	16.83	8.19	13.55	22.40
SD dep. var. control group	7.96	3.87	7.15	10.17	37.42	27.43	34.23	41.70
					Repetition			
Treatment	-4.87***	-3.69	-6.75**	-3.96	-3.80	-1.78	-1.09	-5.16
	(1.79)	(2.58)	(3.01)	(4.22)	(2.37)	(2.23)	(3.13)	(3.42)
Number of observations	277	95	104	78	17276	3629	5490	8157
Number of clusters					277	95	104	78
Mean dep. var. control group	22.84	14.37	29.59	23.61	23.25	12.20	27.72	24.78
SD dep. var. control group	15.27	12.86	14.91	13.71	42.25	32.74	44.77	43.18

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school and student. School-level data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses, respectively. Outcome variables in the panel headers. Robust standard errors for school-level regressions and standard errors clustered at the school level for student-level regressions in parentheses. Coefficients are sample-weighted average treatment effects of the within-block difference-in-means (Blocked DIM). The mean and standard deviation of the dependent variable in the control group is unconditional.

Table E8: Impact on Progression – Interaction-Weighted Estimator (IWE)

	Grade level				*Student level*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
Passing								
Treatment	4.70***	2.39	8.45***	2.51	4.42**	1.29	6.88**	4.15
	(1.74)	(2.45)	(3.15)	(3.43)	(2.15)	(2.67)	(2.86)	(3.98)
Percentage difference between IWE and OLS	0.055	-1.937	-0.165	2.314	-2.060	-0.370	-1.720	-2.376
P-value for joint test of equality between IWE and OLS	0.033	0.272	0.183	0.059	0.003	0.515	0.018	0.045
P-value for joint Wald Test for interactions	0.325	0.589	0.501	0.399	0.111	0.817	0.258	0.253
Dropout								
Treatment	-0.21	-0.15	-1.61	1.57	-0.72	0.27	-4.28**	1.23
	(0.79)	(0.76)	(1.22)	(2.12)	(1.35)	(1.37)	(1.76)	(2.55)
Percentage difference between IWE and OLS	3.861	-5.451	0.015	-1.744	-15.118	2.681	-1.704	9.284
P-value for joint test of equality between IWE and OLS	0.227	0.698	0.186	0.288	0.008	0.297	0.019	0.084
P-value for joint Wald Test for interactions	0.562	0.926	0.482	0.681	0.100	0.605	0.222	0.305
Repetition								
Treatment	-4.49***	-2.24	-6.84**	-4.09	-3.70**	-1.56	-2.60	-5.39*
	(1.63)	(2.29)	(2.79)	(3.49)	(1.62)	(1.88)	(2.55)	(2.88)
Percentage difference between IWE and OLS	-0.117	-1.696	-0.208	0.712	0.984	0.128	-1.747	0.074
P-value for joint test of equality between IWE and OLS	0.249	0.358	0.179	0.485	0.058	0.886	0.013	0.186
P-value for joint Wald Test for interactions	0.599	0.686	0.477	0.832	0.309	0.986	0.147	0.550

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school and student. School-level data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses, respectively. Outcome variables in the panel headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors for school-level regressions and standard errors clustered at the school level for student-level regressions in parentheses. All estimates are IWE as in Gibbons et al. (2018).

Table E9: Impact on Progression – Regression-Weighted Estimator (RWE)

	Grade level				*Student level*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
Passing								
Treatment	4.81***	2.54	8.52***	2.60	4.42**	1.29	6.88**	4.15
	(1.79)	(2.49)	(3.23)	(3.54)	(2.23)	(2.65)	(3.13)	(4.09)
Percentage difference between RWE and OLS	2.417	4.074	0.642	5.752	-1.995	-0.177	-1.681	-2.311
P-value for joint test of equality between RWE and OLS	0.028	0.250	0.423	0.203	0.673	0.924	0.338	0.824
Dropout								
Treatment	-0.25	-0.17	-1.65	1.55	-0.72	0.27	-4.28**	1.23
	(0.81)	(0.77)	(1.24)	(2.17)	(1.38)	(1.37)	(1.82)	(2.59)
Percentage difference between RWE and OLS	22.060	6.700	2.983	-2.997	-14.903	1.933	-1.652	9.126
P-value for joint test of equality between RWE and OLS	0.033	0.579	0.054	0.416	0.282	0.665	0.125	0.687
Repetition								
Treatment	-4.56***	-2.37	-6.86**	-4.15	-3.70**	-1.56	-2.60	-5.39*
	(1.66)	(2.32)	(2.85)	(3.54)	(1.70)	(1.87)	(2.81)	(2.92)
Percentage difference between RWE and OLS	1.529	3.893	0.094	2.299	1.014	0.163	-1.730	0.092
P-value for joint test of equality between RWE and OLS	0.110	0.202	0.913	0.334	0.788	0.888	0.636	0.986

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school and student. School-level data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are constructed from Rio Grande do Norte 2016 and 2017 censuses, respectively. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors for school-level regressions and standard errors clustered at the school level for student-level regressions in parentheses. All estimates are RWE as in Gibbons et al. (2018).

Table E10: Impact on Teacher Retention – Alternative Estimators

	(1) All	(2) 5th	(3) 6th	(4) 10th
Panel A – Blocked Difference-in-Means				
Treatment	0.035	-0.025	0.090**	0.005
	(0.031)	(0.070)	(0.037)	(0.046)
Number of observations	1882	189	784	909
Number of clusters	277	95	104	78
Mean dep. var. control group	0.709	0.761	0.691	0.714
SD dep. var. control group	0.454	0.428	0.463	0.452
Panel B – Interaction-Weighted Estimator (IWE)				
Treatment	0.036	-0.065	0.063*	0.034
	(0.029)	(0.064)	(0.036)	(0.049)
Percentage difference between IWE and OLS	-0.384	0.757	-1.242	2.957
P-value for joint test of equality between IWE and OLS	0.052	0.663	0.009	0.280
P-value for joint Wald Test for interactions	0.296	0.901	0.175	0.638
Panel C – Regression-Weighted Estimator (RWE)				
Treatment	0.036	-0.063	0.063*	0.033
	(0.029)	(0.063)	(0.037)	(0.050)
Percentage difference between RWE and OLS	0.648	-2.153	-0.965	2.650
P-value for joint test of equality between RWE and OLS	0.789	0.784	0.184	0.555

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from Rio Grande do Norte 2016 and 2017 teacher censuses. Unit of observation: teacher. In Panel A, coefficients are sample-weighted average treatment effects of the within-block difference-in-means (Blocked DIM). In Panel B and C, estimates are IWE and RWE, respectively, as in Gibbons et al. (2018). Standard errors clustered at the school level in parentheses.

Table E11: Impact on Socio-Emotional Skills – Controlling for Students' Characteristics

	(1) Agreeableness	(2) Conscientiousness	(3) Extroversion	(4) Neuroticism	(5) Openness
All schools					
Treatment	0.059	0.118*	0.099*	0.098	0.042
	(0.063)	(0.061)	(0.059)	(0.061)	(0.061)
Number of observations	2141	2141	2141	2141	2141
Number of clusters	213	213	213	213	213
Mean dep. var. control group	4.419	4.347	4.206	3.993	4.122
SD dep. var. control group	0.983	1.060	0.788	0.741	0.964
5th grade – Primary schools					
Treatment	-0.058	-0.001	-0.017	-0.023	-0.102
	(0.087)	(0.096)	(0.092)	(0.083)	(0.091)
Number of observations	778	778	778	778	778
Number of clusters	82	82	82	82	82
Mean dep. var. control group	4.552	4.436	4.347	4.099	4.262
SD dep. var. control group	1.034	1.124	0.834	0.754	0.987
6th grade – Lower secondary schools					
Treatment	0.076	0.139	0.163	0.145	0.070
	(0.119)	(0.109)	(0.104)	(0.119)	(0.108)
Number of observations	796	796	796	796	796
Number of clusters	82	82	82	82	82
Mean dep. var. control group	4.356	4.281	4.125	3.917	3.962
SD dep. var. control group	1.077	1.152	0.860	0.752	1.054
10th grade – Upper secondary schools					
Treatment	0.182	0.235**	0.148*	0.187*	0.194*
	(0.110)	(0.097)	(0.086)	(0.101)	(0.105)
Number of observations	567	567	567	567	567
Number of clusters	49	49	49	49	49
Mean dep. var. control group	4.324	4.325	4.130	3.965	4.180
SD dep. var. control group	0.674	0.752	0.519	0.687	0.700

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All regressions are OLS with strata (i.e., region and grade) fixed effects and control for students' characteristics, such as age, gender and race dummies (white, indigenous, black, or *pardo*), whether they receive *Bolsa Família*, and whether they use school transportation. Standard errors clustered at the school level are in parentheses.

Table E12: Impact on Socio-Emotional Skills – Blocked Difference-in-Means

	(1) Agreeableness	(2) Conscientiousness	(3) Extroversion	(4) Neuroticism	(5) Openness
All schools					
Treatment	0.013	0.085	0.107*	0.013	0.019
	(0.059)	(0.061)	(0.063)	(0.050)	(0.059)
Number of observations	3560	3560	3560	3558	3560
Number of clusters	235	235	235	235	235
Mean dep. var. control group	4.413	4.331	4.199	4.007	4.105
SD dep. var. control group	0.975	1.053	0.777	0.738	0.970
5th grade – Primary schools					
Treatment	-0.025	0.072	0.053	0.004	-0.085
	(0.102)	(0.104)	(0.112)	(0.079)	(0.103)
Number of observations	1296	1296	1296	1294	1296
Number of clusters	85	85	85	85	85
Mean dep. var. control group	4.468	4.359	4.287	4.040	4.193
SD dep. var. control group	1.049	1.108	0.851	0.738	0.997
6th grade – Lower secondary schools					
Treatment	0.068	0.170*	0.203**	0.009	0.120
	(0.095)	(0.100)	(0.099)	(0.084)	(0.095)
Number of observations	1270	1270	1270	1270	1270
Number of clusters	87	87	87	87	87
Mean dep. var. control group	4.390	4.265	4.156	3.971	3.950
SD dep. var. control group	1.090	1.176	0.858	0.770	1.089
10th grade – Upper secondary schools					
Treatment	0.011	-0.006	0.075	0.036	0.072
	(0.098)	(0.086)	(0.091)	(0.091)	(0.083)
Number of observations	994	994	994	994	994
Number of clusters	63	63	63	63	63
Mean dep. var. control group	4.378	4.387	4.152	4.017	4.212
SD dep. var. control group	0.663	0.761	0.514	0.692	0.701

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. Coefficients are sample-weighted average treatment effects of the within-block difference-in-means (Blocked DIM). Standard errors clustered at the school level are in parentheses. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score.

Table E13: Impact on Socio-Emotional Skills – Interaction-Weighted Estimator (IWE)

	(1) Agreeableness	(2) Conscientiousness	(3) Extroversion	(4) Neuroticism	(5) Openness
All schools					
Treatment	0.046	0.113**	0.113**	0.036	0.055
	(0.054)	(0.052)	(0.053)	(0.046)	(0.052)
Percentage difference between IWE and OLS	-4.587	-1.692	-2.212	-1.581	-5.082
P-value for joint test of equality between IWE and OLS	0.221	0.002	0.101	0.211	0.059
P-value for joint Wald Test for interactions	0.462	0.122	0.483	0.680	0.303
5th grade – Primary schools					
Treatment	0.023	0.093	0.048	-0.020	-0.062
	(0.095)	(0.094)	(0.096)	(0.073)	(0.092)
Percentage difference between IWE and OLS	-0.984	-0.341	-1.650	2.051	1.703
P-value for joint test of equality between IWE and OLS	0.745	0.890	0.467	0.868	0.861
P-value for joint Wald Test for interactions	0.945	0.989	0.815	0.984	0.980
6th grade – Lower secondary schools					
Treatment	0.078	0.173*	0.206**	0.060	0.138
	(0.097)	(0.096)	(0.096)	(0.091)	(0.096)
Percentage difference between IWE and OLS	-0.020	-0.147	-0.561	2.750	-0.233
P-value for joint test of equality between IWE and OLS	0.211	0.487	0.377	0.621	0.308
P-value for joint Wald Test for interactions	0.517	0.795	0.703	0.886	0.625
10th grade – Upper secondary schools					
Treatment	0.035	0.063	0.080	0.079	0.103
	(0.083)	(0.068)	(0.074)	(0.070)	(0.071)
Percentage difference between IWE and OLS	-16.644	-7.669	-5.785	-3.637	-6.341
P-value for joint test of equality between IWE and OLS	0.051	0.000	0.075	0.019	0.007
P-value for joint Wald Test for interactions	0.223	0.007	0.384	0.172	0.134

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All estimates are IWE as in Gibbons et al. (2018). Standard errors clustered at the school level are in parentheses. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score.

Table E14: Impact on Socio-Emotional Skills – Regression-Weighted Estimator (RWE)

	(1) Agreeableness	(2) Conscientiousness	(3) Extroversion	(4) Neuroticism	(5) Openness
All schools					
Treatment	0.046	0.113**	0.114**	0.036	0.055
	(0.056)	(0.054)	(0.054)	(0.047)	(0.054)
Percentage difference between RWE and OLS	-4.761	-1.763	-2.109	-1.953	-5.274
P-value for joint test of equality between RWE and OLS	0.248	0.183	0.109	0.587	0.053
5th grade – Primary schools					
Treatment	0.023	0.093	0.048	-0.020	-0.062
	(0.096)	(0.095)	(0.096)	(0.074)	(0.093)
Percentage difference between RWE and OLS	-1.707	-0.373	-1.363	1.391	1.820
P-value for joint test of equality between RWE and OLS	0.889	0.895	0.796	0.907	0.650
6th grade – Lower secondary schools					
Treatment	0.078	0.173*	0.207**	0.060	0.138
	(0.099)	(0.098)	(0.097)	(0.092)	(0.098)
Percentage difference between RWE and OLS	0.026	-0.111	-0.505	2.322	-0.220
P-value for joint test of equality between RWE and OLS	0.992	0.925	0.636	0.546	0.874
10th grade – Upper secondary schools					
Treatment	0.035	0.063	0.080	0.079	0.102
	(0.092)	(0.079)	(0.078)	(0.074)	(0.080)
Percentage difference between RWE and OLS	-16.960	-8.219	-5.698	-3.997	-6.596
P-value for joint test of equality between RWE and OLS	0.175	0.058	0.123	0.177	0.080

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: student. All estimates are RWE as in Gibbons et al. (2018). Standard errors clustered at the school level are in parentheses. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score.

Table E15: Impact on Socio-Emotional Skills – School Level Regressions

	(1) Agreeableness	(2) Conscientiousness	(3) Extroversion	(4) Neuroticism	(5) Openness
All schools					
Treatment	0.075	0.142**	0.103	0.087	0.078
	(0.076)	(0.069)	(0.069)	(0.062)	(0.069)
Number of observations	235	235	235	235	235
Mean dep. var. control group	4.379	4.293	4.162	3.966	4.049
SD dep. var. control group	0.566	0.612	0.468	0.355	0.582
5th grade – Primary schools					
Treatment	0.086	0.157	0.050	0.092	-0.010
	(0.125)	(0.120)	(0.118)	(0.102)	(0.121)
Number of observations	85	85	85	85	85
Mean dep. var. control group	4.435	4.323	4.258	3.984	4.162
SD dep. var. control group	0.689	0.651	0.484	0.347	0.562
6th grade – Lower secondary schools					
Treatment	0.098	0.204*	0.274**	0.096	0.182
	(0.120)	(0.111)	(0.124)	(0.096)	(0.124)
Number of observations	87	87	87	87	87
Mean dep. var. control group	4.359	4.211	4.080	3.938	3.853
SD dep. var. control group	0.595	0.718	0.573	0.406	0.702
10th grade – Upper secondary schools					
Treatment	0.039	0.064	0.022	0.071	0.107
	(0.145)	(0.122)	(0.109)	(0.124)	(0.106)
Number of observations	63	63	63	63	63
Mean dep. var. control group	4.330	4.372	4.149	3.984	4.182
SD dep. var. control group	0.262	0.315	0.161	0.285	0.258

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school. All regressions are OLS with strata (i.e., region and grade) fixed effects and analytic weights for the number of students enrolled in the grade of interest. Robust standard errors are in parentheses. The coefficients are expressed in terms of standard deviations from the control group, while the mean and standard deviation of the dependent variable refer to the raw values in the control group. 'Neuroticism' is reverse-coded so that a positive coefficient implies a lower level of neuroticism score.

F Back-of-the-Envelope Estimations

F.1 IDEB

Methodology

The Brazilian Education Development Index (*Índice de Desenvolvimento da Educação Básica* – IDEB) was created by the INEP in 2007 as an indicator that aggregates the two main drivers of education quality: student proficiency as quantified by standardized exams and student attainment as measured by grade passing rates.[51] Since then, IDEB has been regularly employed to monitor the evolution of Brazilian education system and to compare different state experiences.

In order to have a comparable measure of education learning gains, IDEB uses the national standardized exams in math and Portuguese, known as SAEB. This test is administered to all public and private schools every second year. In particular, students in the last year of primary, lower- and upper-secondary schools, i.e., 5th, 9th, and 12th grades, are evaluated. SAEB tests are based on IRT so as to define a unique scale for all grades and years of the national education system. This is done by including items from the previous grades and years in the test.[52]

To compute IDEB, SAEB scores are standardized in a scale between 0 and 10, following the equation

$$N_{sj} = \frac{score_{sj} - min_j}{max_j - min_j} \cdot 10 \tag{F1}$$

where j is the subject of the test, i.e., either math or Portuguese, and s is the school identifier. min_j and max_j are the inferior and superior limits of subject j in the 1997 SAEB (the first year in which the test was administered nationwide). Namely, these

[51] You can find the informative and technical notes (in Portuguese) on how MEC compiles IDEB at http://download.inep.gov.br/educacao_basica/portal_ideb/o_que_e_o_ideb/nota_informativa_ideb.pdf or http://download.inep.gov.br/educacao_basica/portal_ideb/o_que_e_o_ideb/Nota_Tecnica_n1_concepcaoIDEB.pdf. Our methodological discussion faithfully reflects the contents of these two documents.

[52] Besides the test, students, teachers, and principals are subject to socio-economic and cultural questionnaires, which are used by the MEC to foster the understanding of the tested schools.

limits were computed by taking the values 3 SDs, σ_j, away from the average, μ_j, of the 1997 scores in each discipline

$$min_j = \mu_j - 3 \cdot \sigma_j \, ; \quad max_j = \mu_j + 3 \cdot \sigma_j \tag{F2}$$

Finally, the arithmetic mean of math and Portuguese standardized scores is taken

$$N_s = \frac{N_{s,j=math} + N_{s,j=Portuguese}}{2} \tag{F3}$$

With regard to student attainment, IDEB uses an indicator of achievement at the school level, P_s, which is obtained by taking the inverse of the average of the passing rates of primary, middle, or high school, T_s. In mathematical notation,

$$T_s = \frac{\sum_{y=1}^{Y} \frac{1}{p_{sy}}}{Y} \tag{F4}$$

$$P_s = \frac{1}{T_s} \tag{F5}$$

where y is the grade of interest, Y is the total number of grades with positive passing rates in the school s, and p_{sy} is the grade-level passing rate. In the absence of dropout, T_s measures the duration time of a certain stage of education for an average student in school s.

Hence, IDEB results from multiplying the two indicators defined in Equation F3 and F5

$$IDEB_s = N_s \cdot P_s \tag{F6}$$

$$0 \leq N_s \leq 10 \, ; \; 0 \leq P_s \leq 1 \, ; \; 0 \leq IDEB_s \leq 10 \tag{F7}$$

and is equal to the standardized 0-10 score in SAEB adjusted for the average time (in years) it takes to conclude one grade in that stage of education.

Estimation

As mentioned in Section 3.2, the state standardized tests on which we base our analysis were rescaled to SAEB ITR range allowing one to compute N_s, as defined in Equation F3, for each school in our sample. As we described in the paper, the PIP was implemented in the last year of primary school, i.e., the 5th grade, but not in the last years of lower- and upper-secondary school. This means that we are not able to compute the IDEB for those grades, but we focus on the grade of the intervention.

On the other hand, we use passing rates in the grade of the intervention to calculate P_s. Again, as we are looking only at one grade of a stage of education, P_s will be equal to the passing rate (in percentage points) in that grade.

Combining these two variables, we calculate a grade-level measure of IDEB for schools in the treatment and control groups. Therefore, we use this index to estimate the ITT in terms of IDEB points. Namely, we employ the model defined in Equation 1. The results are shown in Table F1.

In line with the baseline results on standardized test scores and passing rates, the only significant effect is found in 6th grade. The intervention had an ITT of 0.28 IDEB points. We take this coefficient to assess how the PIP would move RN across the nation distribution. In order to do so, we compare lower-secondary IDEB in 2015 for all Brazilian states.[53] As one can see in Figure F2, RN was the third worst state in terms of quality of education, after Sergipe and Alagoas. The increase in IDEB caused by PIP, as estimated above, would shift RN from the bottom decile to the third decile according to ITT estimates.[54]

[53] As SAEB tests take place every two years, we are not able to have comparable data from 2016, which was the year in which PIP was actually implemented.

[54] The results are robust to the inclusion of school-level control variables.

Table F1: Impact on IDEB

	(1) 5th	(2) 6th	(3) 10th
Treatment	-0.099	0.282**	0.167
	(0.208)	(0.131)	(0.130)
Number of observations	95	104	78
Mean dep. var. control group	3.606	1.649	2.062
SD dep. var. control group	1.056	0.696	0.572

Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Unit of observation: school. Regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors are in parentheses.

F.2 Net Present Value of Increased Learning

Increased learning is associated with long-term labor market returns, assuming that the accumulation of human capital is sustained over time. In this subsection, we follow the method proposed by Evans and Yuan (2019) to translate the impact of the education intervention in net present value (NPV) of potential increased lifetime earnings. The NPV is defined as

$$NPV = \sum_{k=20-\alpha}^{N} \frac{\Delta Y \cdot \beta \cdot w}{(1+i)^k} \tag{F8}$$

where ΔY is the number of equivalent years of education caused by the intervention, β is the return to one year of education, w is the real wage, i is the discount rate, α is the age at which the student was targeted by the intervention, and N is his/her expected work life.

Hence, $\Delta Y \cdot \beta$ represents the predicted wage increase, stemming from the learning improvement. Assuming constant wages over time, this translates into an additional income of $\Delta Y \cdot \beta \cdot w$ for an average worker.[55] As students enter the labor market only in a later stage (when they are 20 years old), these wage gains are discounted by $k = 20 - \alpha$ years. Therefore, we sum the yearly increases in NPV across the whole work life of a student.

We use the 2016 Annual Social Information Report (*Relação Anual de Informações Soci-*

[55]This is a conservative approach: as we expect wages to grow over time, the actual NPV from the intervention may be higher than the one we estimate hereafter.

ais – RAIS) from the Ministry of Labor and Employment to retrieve the average wage in RN (this refers to the formal sector) and, therefore, to estimate the return to education in RN through a conventional Mincerian equation (Mincer, 1974). Namely, the average wage in 2016 was BRL 24,486.48, i.e., around USD 6,000. In line with recent estimates by Psacharopoulos and Patrinos (2018), we find the return to one extra year of education in RN to be around 10 percent. The age of 6th graders, who received the intervention, was on average 12 years, and we assume the expected work life to be 40 years (which means an average worker retires when he/she is 60). Finally, the discount rate is taken at 3 percent.

Using our ITT estimates, as computed in Section 4.2, we find that PIP would increase annual wages by 5 percent. This would mean a shift of the median worker to the 6th, or 7th, decile, respectively, in the wage distribution of RN (Figure F3). Considering the whole work life, the intervention has a predicted NPV between BRL 29,148.97 and 52,468.15, i.e., USD 7,287.24 to 13,117.03. This is equivalent to about one average annual Brazil income per capita.

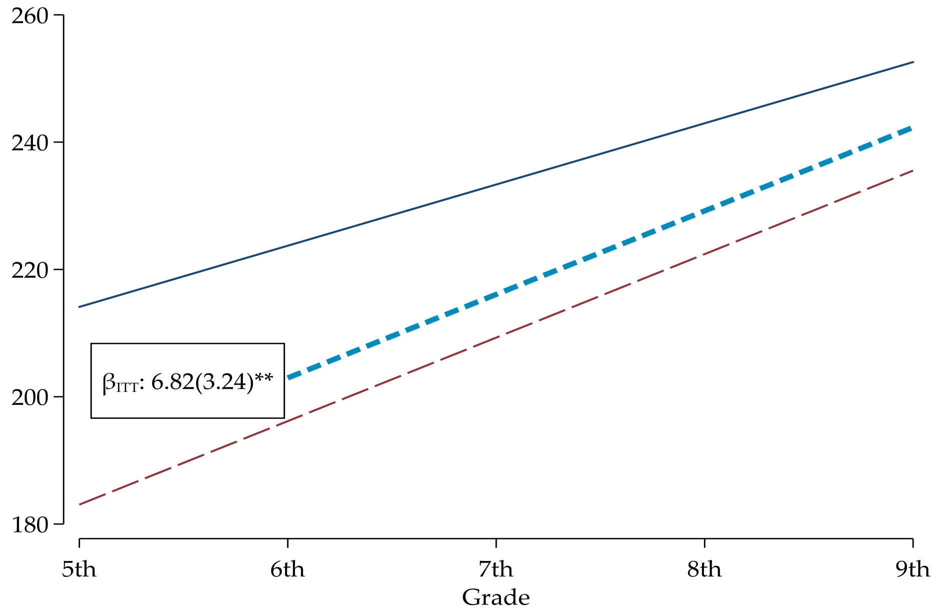
Table F2: Impact on Teacher Retention – Heterogeneity by Qualification

	(1) All	(2) 5th	(3) 6th	(4) 10th
Treatment	0.039	-0.046	0.062	0.044
	(0.031)	(0.082)	(0.042)	(0.050)
Treatment × Teacher has specialization and/or master	-0.009	-0.053	0.014	-0.032
	(0.043)	(0.135)	(0.064)	(0.062)
Teacher has specialization and/or master	-0.030	0.003	-0.078	0.010
	(0.028)	(0.086)	(0.048)	(0.033)
Constant	0.725***	0.767***	0.722***	0.716***
	(0.021)	(0.050)	(0.031)	(0.031)
Total effect on teachers with specialization and/or master: Treatment + Treatment × master dummy				
$\sum \hat{\beta}$	0.031	-0.099	0.076	0.013
P-value	0.464	0.357	0.183	0.852
Unconditional mean of the dependent variable in the control group:				
Teachers without specialization and/or master	0.693	0.756	0.644	0.725
Teachers with specialization and/or master	0.720	0.765	0.722	0.706
Number of observations	1882	189	784	909
Number of clusters	277	95	104	78

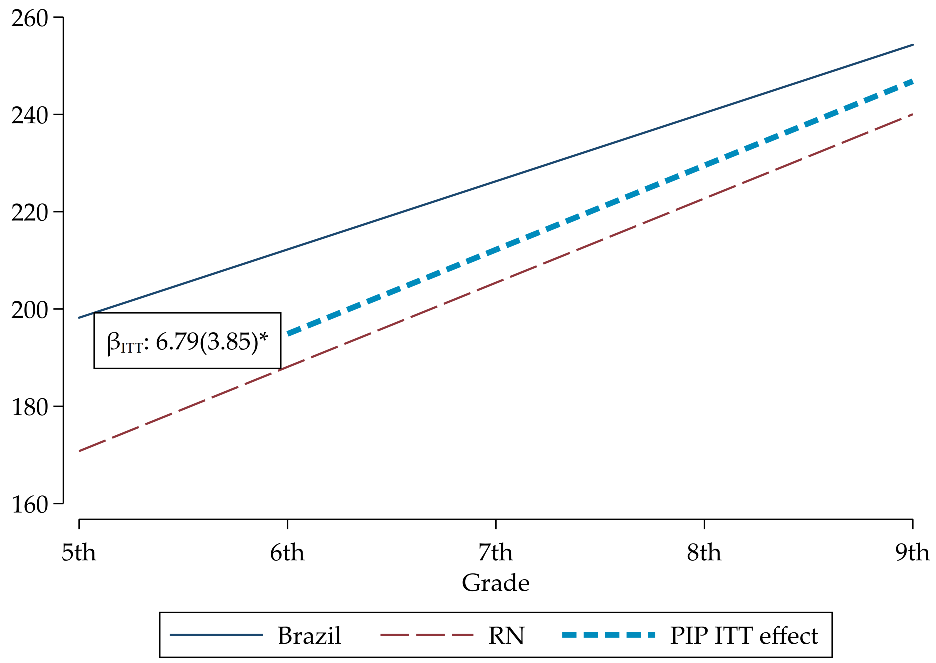
Note: *Significant at 10 percent. **Significant at 5 percent. ***Significant at 1 percent. Data are from *Sistema Integrado de Gesto da Educao* (SIGEduc) and student-level data are from Rio Grande do Norte census. Unit of observation: school-grade. Outcome variables in the column headers. All regressions are OLS with strata (i.e., region and grade) fixed effects. Robust standard errors in parentheses. Sample: schools treated at 6th grade. All outcome variables (in the panel headers) are dummy variables and regressions are linear probability model with strata (i.e., region) fixed effects. Standard errors clustered at the school level in parentheses.

Figure F1: Learning Gains in 6th Grade Rescaled to SAEB – Projection over Time

(a) Math

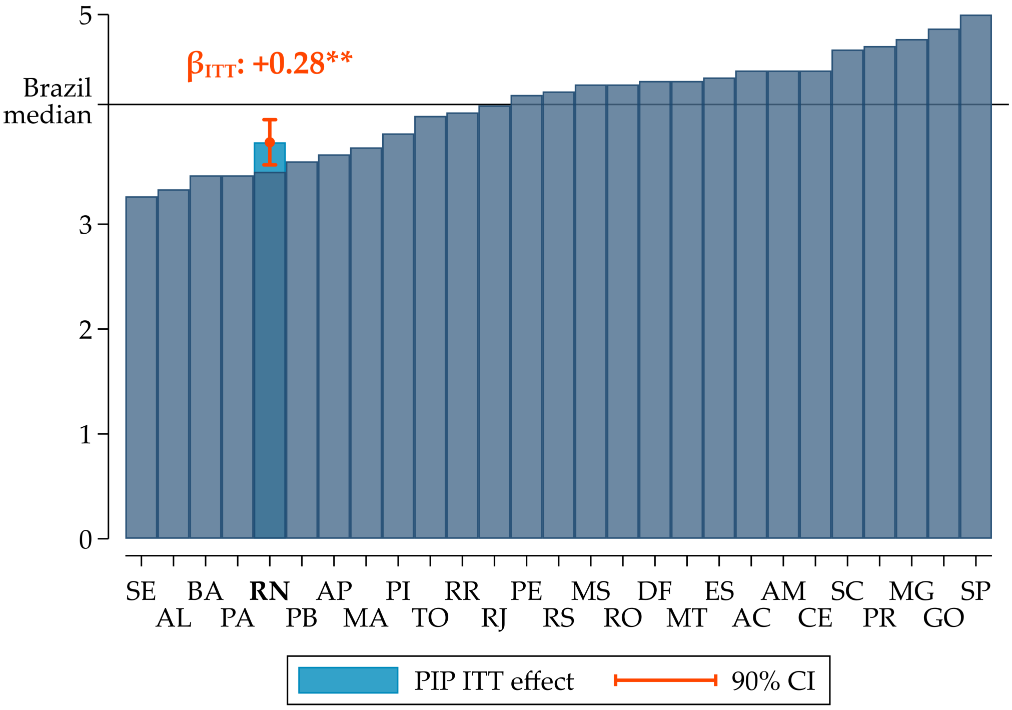


(b) Portuguese



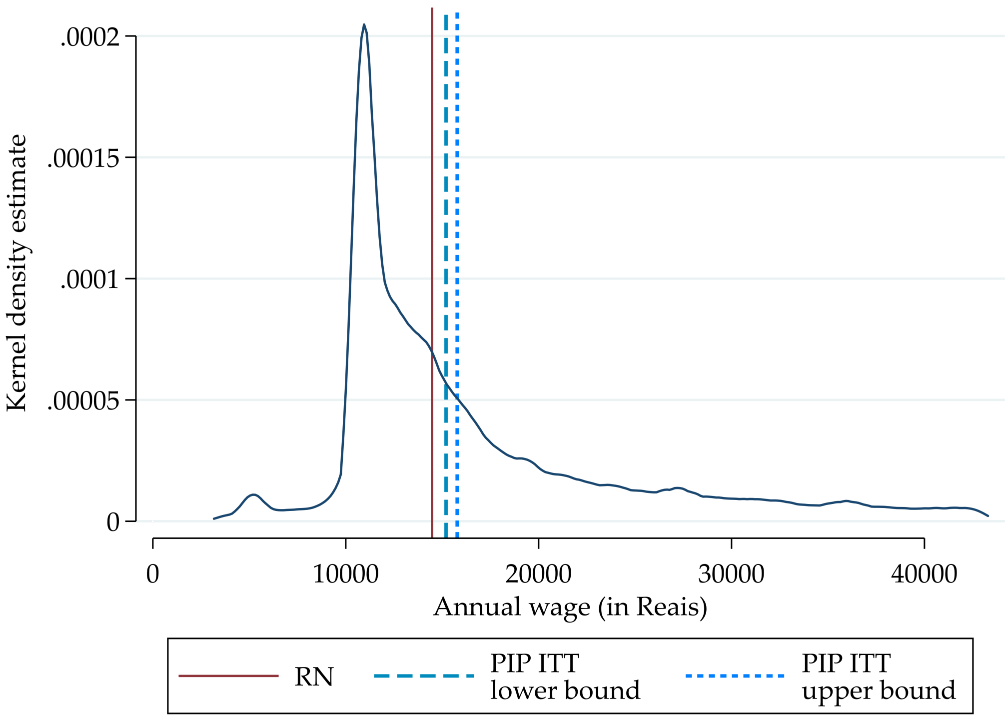
Note: We use data from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) for state public schools in Rio Grande do Norte and Brazil. In particular, we use the average for the cohort who was in 5th grade in 2013 and 9th grade in 2017. The points in 6th, 7th, and 8th grades are linear interpolation. The PIP intent-to-treat effect on 6th graders is estimated through OLS with strata (i.e., region and grade) fixed effects and standard errors clustered at the school level. **, and * indicate significance at the 5, and 10 percent critical level.

Figure F2: Learning Gains in 6th Grade Rescaled to *IDEB* – Comparison with Other Brazilian States



Note: We use data from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) for state public schools. The bars show the average IDEB by state in 2015. The PIP intent-to-treat effect on 6th graders is estimated through OLS with strata (i.e., region and grade) fixed effects and robust standard errors. ** indicate significance at the 5 percent critical level. See Section F.1 for the methodology we follow to compute IDEB for our grades of interest.

Figure F3: Learning Gains in 6th Grade Rescaled to Annual Wage



Note: Kernel densities are computed using Epanechnikov kernel function. The three horizontal lines represent the median wage of Rio Grande do Norte, which is considered as counterfactual, and of the median PIP student, assuming the effects on equivalent years of education estimated in Section 4.2 through an OLS model. The sample is the universe of formal workers in Rio Grande do Norte in 2016. Data are from *Relao Anual de Informaes Sociais* (RAIS). N = 801,956.

Online Appendix References

EVANS, D. K. AND F. YUAN (2019): "Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms," Policy Research Working Paper No. 8752, The World Bank: Washington, DC.

GIBBONS, C. E., J. C. S. SERRATO, AND M. B. URBANCIC (2018): "Broken or Fixed Effects?" *Journal of Econometric Methods*, 8.

IMBENS, G. W. AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

LEE, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 76, 1071–1102.

MINCER, J. (1974): *Schooling, Experience, and Earnings*, National Bureau of Economic Research: Cambridge, MA.

MONTIEL OLEA, J. L. AND C. PFLUEGER (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

PSACHAROPOULOS, G. AND H. A. PATRINOS (2018): "Returns to Investment in Education: A Decennial Review of the Global Literature." *Education Economics*, 26, 445–458.