Supporting Teacher Autonomy to Improve Education Outcomes: Experimental Evidence from Brazil*

Caio Piza† Astrid Zwager† Matteo Ruzzante† Rafael Dantas†
Andre Loureiro‡

Abstract

What is the impact of greater teacher autonomy on student learning? We provide experimental evidence of a program in Brazil that supported teachers, through a combination of technical assistance and a small grant, to autonomously develop and implement an innovative project aimed at engaging their students. We find that the program improved student learning by 0.15 SD and grade passing by 13% in 6^{th} grade, a critical year of transition from primary to lower-secondary education. We explore two mechanisms: teacher turnover and student socio-emotional skills. Teacher turnover is reduced by 20.7% and impacts on student outcomes are concentrated in those schools with the largest reduction. We also find positive impacts on conscientiousness and extroversion among students. The results suggest that, increasing autonomy of public servants can improve service delivery, even in a low capacity context.

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

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 São Paulo School of Economics (EESP-FGV). 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 State Secretariat of Education for being great research partners. Computational reproducibility was verified by DIME Analytics. Details of the reproducibility checklist can be found in the Online Appendix. The reproducibility package is available on GitHub. 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.

†Development Impact Evaluation (DIME), The World Bank, 1818 H Street NW, Washington, DC 20433, United States.

‡Education Global Practice, The World Bank.

1 Introduction

Over the last three decades many countries have succeeded in putting kids in school, but gains in learning have been limited (World Bank, 2018). Improving quality 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, 2012; Chetty et al., 2014b). Despite increasing resource allocation to the sector, governments have struggled to substantially improve education outcomes (McEwan, 2015; Glewwe and Muralidharan, 2016). The recent World Development Report (WDR) points to a 'learning crisis' faced by many countries, including Brazil, 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., 2014a,b; Araujo et al., 2016; Jackson, 2018). While policies tend to put more emphasis on monetary incentives (e.g., salary increase and performance-based payments) or on enhancing qualifications (Evans and Popova, 2016), the role of teacher motivation is often neglected (World Bank, 2018). We present experimental evidence of an education policy in Brazil that seeks to motivate teachers by providing them with the autonomy to design and implement a local project to tackle their specific issues instead of "prescribing solutions".[1]

The effect of assigning civil servants with more autonomy is an empirical question. On the one hand, increasing local autonomy may encourage agents to reduce effort due to limited ability of central government to observe and reward effort accordingly. For example, decentralization may backfire if resources are captured by local entities or used inefficiently (Burgess et al., 2012; Banerjee et al., forthcoming). On the other hand, greater autonomy can improve service delivery by providing a non-monetary incentive for agents and add meaning to the job (Cassar and Meier, 2018)[2] and by leveraging their superior knowledge of local context (Duflo et al., 2018; Rogger and Somani, 2018). Rasul et al. (2018) and Rasul and Rogger (2018) find that more autonomy is correlated with quality and completion of public projects delivered even in contexts of low

[1] The paper does not speak to the wide literature on school decentralization, which involves allowing local management of resources and/or curriculum (Hanushek et al., 2013).

[2] The association between autonomy and motivation is at the foundation of Self-Determination Theory in the social psychology literature (Deci and Ryan, 1985; Ryan and Deci, 2017). Studies have focused on how monetary rewards might crowd out motivation, as they undermine autonomous decision-making (Deci, 1971; Amabile et al., 1976; Pritchard et al., 1977) and how non-monetary incentives, that give greater autonomy, can enhance motivation (Zuckerman et al., 1978)

government capacity, and Bandiera et al. (2020) suggests autonomy can reduce misalignment of incentives between officials and taxpayers with potential welfare benefits for society. We investigate whether increasing autonomy of teachers can achieve efficiency gains in the delivery of public education.

We study the Pedagogical Innovation Project (*Projeto de Inovação Pedagógica* – PIP), which was implemented by the State Secretariat of Education (SEE) of Rio Grande do Norte (RN). RN consistently scores at the bottom of the Brazilian Education Development Index (*Índice de Desenvolvimento da Educação Básica* – IDEB).[3] Through seminars and the support from a dedicated mentor, teachers developed a diagnostic of their main pedagogical challenges and a context-specific project to address them. Mentors complemented local capacity while ensuring close ties with central government, possibly reducing moral hazard concerns associated with strategic behavior of local staff. Approved proposals were awarded financial support to implement the projects, ranging from about US$ 7,500 to US$ 11,000[4], or median 139 US$ per student, i.e., 3.6% of average annual expenditure per student in Brazil (OECD, 2016). The decentralized approach sought to ensure relevance of the interventions as well as motivate teachers by giving them the autonomy over design and implementation.

Our experiment focuses on the 2016 edition, 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). Out of 299 schools eligible to receive the project in 2016, 130 schools were randomly invited to participate and submit a proposal.

We show that the project had substantial impacts on both learning and progression for 6th grade students, a critical grade in school transition from primary to lower-secondary education (Santos et al., 2017). Our ITT estimates point to an impact of 0.18 SD on math and 0.16 SD in Portuguese and slightly lower impacts on human (0.10 SD) and natural sciences (0.12 SD). We estimate the average impact on learning to be the equivalent to 0.5 extra year of schooling, or 0.36 years per US$ 100 spent. Consistent with the results on learning gains, we find substantial

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

[4]Equivalent to 30,000 to 45,000 Brazilian Reais, using exchange rate on 12/31/2015.

improvements in Grade 6 passing rates, which are estimated to increase by 8.5 percentage points (pp), a 13% improvement compared to the control mean. A back-of-the-envelop calculation of the combined effect of increased learning and higher probability of finishing high school suggests a net present value of the expected years of schooling on future earnings ranging between 7 to 13 thousand US dollars or 28 to 52 Brazilian minimum wages. Compared with the cost of the program per student (US$ 139), the estimated NPV suggests that PIP was a high-return investment for the state.

We empirically investigate two, potentially complementary, channels through which the program may have impacted learning and grade progression: (1) teacher turnover and (2) student socio-emotional skills.

We hypothesize that turnover drops if teachers feel more motivated/committed to implement their own pedagogical projects during the academic year. A drop in teacher turnover is in turn expected to affect student outcomes based on the well documented negative relationship between high teacher turnover and learning (Akhtari et al., 2018; Ronfeldt et al., 2013; Jackson et al., 2014). We estimate that the PIP reduced teacher turnover in Grade 6 by 20.7% over the control mean. To test whether the reduction in teacher turnover affected final student outcomes, we estimate heterogeneous effects by teacher turnover at baseline. We find that the impacts on both teacher turnover and learning are concentrated in schools with higher teacher turnover at baseline. Impacts of the program in schools with high teacher turnover at baseline approach 0.28 SD on learning and 7 pp on dropout. To assess whether the results are driven by a mechanical reduction in teacher turnover alone, we leverage the fact that most Grade 6 teachers also teach Grade 7. We estimate a similar reduction on 7$^{\text{th}}$ grade teacher turnover, yet do not find impacts on progression rates of 7$^{\text{th}}$ grade students. These results provide suggestive evidence that the increased motivation may not necessarily spill over to other grades in absence of the project.

We also expect the program to impact students' socio-emotional skills either directly by improving student-teacher interactions and boosting students' motivation through the implementation of the innovative projects, or indirectly through impacts on cognitive skills (Cunha and Heckman, 2007, 2008; Cunha et al., 2010). We measure the Big Five personality traits to test this mechanism. We find positive impacts of 0.17 SD on conscientiousness, the trait most commonly associated with the acquisition of cognitive skills (Poropat, 2009; Ivcevic and Brackett, 2014), and 0.20 SD on extroversion for 6^{th} graders. Grade 6 is a critical moment for students as they transition from primary to lower-secondary education. During this transition, students move from having a single teacher to multiple teachers resulting in weaker ties between students and teachers, which has been shown to affect learning and socio-emotional skills (Bedard and Do, 2005; Hanewald, 2013; Santos et al., 2017). Improving teacher and student motivation might counterbalance the weakening of student-teacher interaction at this stage.

Our paper makes three main contributions. First, it provides experimental evidence that increase in the autonomy of local civil service providers, complemented with technical assistance, can improve outcomes of interest even in a low capacity environment. While monetary incentives, such as performance-based payments, have achieved positive results in some contexts (Lavy, 2009; Muralidharan and Sundararaman, 2011; Duflo et al., 2012; Mbiti et al., 2019), these schemes may not be feasible in many developing countries. Our findings suggest that results can be achieved in a more cost effective way by exploiting the potential complementarity between non-monetary incentives and intrinsic motivation (Bowles and Polania-Reyes, 2012; Ashraf et al., 2014). However, complementing local capacity through technical assistance may be critical, as autonomy alone has had limited success in the Brazilian context (Almeida et al., 2016; Oliveira et al., 2016).

Second, while the literature on school grants (Das et al., 2013; Blimpo et al., 2015; Beasley and Huillery, 2017; Carneiro et al., 2020) shows that decentralizing resources and the authority over their use may be a more efficient way to target resources by leveraging better information of local decision-makers, our results show that efficiency gains can be obtained by leveraging mostly existing resources. This result speaks to the growing evidence pointing to the importance of school management in improving learning outcomes through better resource allocation (Abdulkadiroğlu et al., 2011; Dobbie and Fryer, 2013; Rockoff et al., 2012; Taylor and Tyler, 2012; Fryer, 2014; Fryer, 2017).

Finally, there has been increasing attention on building students' socio-emotional skills early in school due to their role in complementing cognitive skills in predicting academic achievement and labor market outcomes (Heckman and Rubinstein, 2001; Heckman et al., 2006; Almlund et al., 2011; Lindqvist and Vestman, 2011). However, it is not common practice to measure socio-emotional skills, especially outside the context of interventions that aim to affect them directly (Heckman, 2000, Heckman and Kautz, 2012, Sanchez Puerta et al., 2016). Our results show that measuring socio-emotional skills can expand the understanding of how these skills can be affected and may avoid understating important welfare benefits of education programs.

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 results, while Section 5 explores potential mechanisms driving the main impacts. Section 6 provides back-of-the-envelope estimates for the impact of the program on school quality indicators and individuals' expected earnings. Section 7 concludes with policy recommendations.

2 Context and Intervention

2.1 Education in Brazil and Rio Grande do Norte

Brazil has made significant progress to guarantee universal access to primary education, reaching a 99% enrollment rate for children between the age of 6 and 14 years in 2018.[5] Despite this substantial progress, large challenges remain to keep kids in school and ensure quality of education. Grade repetition and dropout rates in primary and secondary schools are among the highest in the world. Large age-grade distortions are found across all grades and an average student spends 15 years – instead of 12 – to graduate from high school. Among the youth of 19 years old, only 63.5% have graduated high school.[6] Despite the largest improvements in math score in the PISA evaluation between 2003 and 2012, Brazil still ranks below all LAC countries except for Peru and the Dominican Republic (OECD, 2015).

A major constraint to school quality and student achievement in Brazil is principal and teacher turnover (Akhtari et al., 2018). In the RN public school system, thirty percent of teachers leave their schools each year, potentially disrupting school operations and compromising personnel collaboration.[7] Using school-level data from INEP, we find that teacher permanence is positively correlated with passing rates and negatively correlated with age-grade distortion, retention and dropout rates, for both primary and secondary schools (Table A1).[8]

[5]Source: National Household Survey – *Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua* – and School Census.

[6]Source: *PNAD Contínua* and School Census.

[7]Source: teacher census (INEP). One reason for the high turnover relates to how placement of teachers is organized in Brazil. Teachers are initially placed at any school with a vacancy, with limited consideration of their location preferences. Every year, teachers are allowed to compete for new vacancies (Akhtari et al., 2018).

[8]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 5-period time frame, weighting for the number of teachers in a school. The

These national figures hide a high degree of regional variation (Figure A1). In this paper we study an education project implemented by the RN state government, one of Brazil's poorest states. In the 2015 national standardized exam[9] RN state schools scored at the bottom of the learning distribution in both primary and lower secondary education.[10] The difference in 5$^{\text{th}}$ grade proficiency levels between the average student in RN and the best performing state is the equivalent of 2.5 years of education.[11] The low level of learning is reflected in the state's progression indicators. In 2015, average school dropout at upper secondary education was 12.4% compared to the national average of 8.8%.[12] The combination of high dropout rates and low learning outcomes put RN state schools near the bottom of the national quality of education indicator (Figure A1).

2.2 The Pedagogical Innovation Project (PIP)

The Pedagogical Innovation Project (*Projeto de Inovação Pedagógica* – PIP) was developed by the RN SEE and aimed at improving both student progression and learning outcomes in primary and secondary state schools, which represent 16% of elementary schools, 41% of middle schools, and 94% of high schools in the public education system. The program targets Grades 4, 5, 6 and 10, the grades with the most critical dropout and retention rates (Figure A2). The intervention has three main components: i) High degree of autonomy for teachers to design and implement a project based on a context-specific diagnostic of the main challenges, with the SEE having only an advisory role to assure minimum quality standards; ii) Continuous technical support to schools for the design and implementation of the project; iii) A grant to implement the project.

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

index ranges from 0 to 5, where a higher number indicates more regularity of the teacher pool in a school.

[9](*Sistema de Avaliação da Educação Básica* – SAEB)

[10]2015 is the year prior to the roll-out of the interventions we study in this paper.

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

[12]Source: INEP.

The PIP was launched in 2014 and between the 2015 to 2018 school years covered a total of 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.

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 to their context. Each school is provided an individualized report card comparing their test scores and passing grades with average of the state, region, and their city.

Following the workshop, each school is assigned a mentor (*professor orientador*), who is part of the SEE central team.[13] Each mentor is assigned to 10 schools on average. The mentor first works with the school to prepare a diagnostic of their challenges, such as low academic performance, grade retention, indiscipline, lack of motivation, or school dropout. Based on the diagnostic, schools 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.

Implementation Support and Monitoring

Schools with approved proposals are awarded with a fixed amount of funding to execute their projects. Schools can only spend the 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 R$ 30,000 to 45,000, i.e., US$ 7,576 to 11,364 (Figure A3).[14] The median transfer per enrolled student was R$ 555.55, the equivalent of US$ 139, which represents about 3.6% 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.

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

[14] Using exchange rate on 12/31/2015.

Characteristics of Sub-Projects

Schools were encouraged to explore 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. In practice, all submitted proposals were approved. Most proposals fell into one of the following three categories:

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

Communication, media and culture: The focus of this type of project was to introduce students to digital tools and give teachers the opportunity to use new technologies and social media. Examples include the development of videogames and robotics classes.

Culture and arts: The goal of these projects was to explore different forms of cultural and artistic expressions, such as painting, graffiti, dance, theater, cinema, and music.

3 Experimental Design and Data

The PIP was launched in 2014 with implementation taking place in the 2015 school year. Each year a subset of state schools joined the project. Our study focuses on the cohort of schools that initiated design in 2015 and implemented the project in the 2016 school year. 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 were selected to participate each year. To determine the pool of eligible schools for that year, three filters were applied. First, only schools that would not change director between the 2015 and 2016 school year were included to ensure buy-in for the prepared projects. State legislation requires directors to change schools every two years, resulting in about half the schools changing director every year.[15] Second, the 2016 edition targeted the final grade of primary education (5^{th} grade), the first grade of lower secondary education (6^{th} grade) and the first grade of upper secondary education (10^{th} grade).[16] Only schools offering at least one of those three grades were considered. Finally, schools that participate in the Federal project ProEMI (*Ensino Médio Inovador)* were excluded.[17] Out of 639 state schools, 299 were eligible to receive the PIP project in 2016.

The selection of participating schools was done randomly, 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 typically 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^{th}, 6^{th} and 10^{th}) are assigned to participate with the highest target grade they offer. The state is divided in 4 regions and, with the 3 grade levels, this resulted in a total of 12 strata. In each stratum, around 40% of the schools were allocated to the treatment group. In each selected school only the highest target grade, i.e., 5^{th}, 6^{th}, or 10^{th}, is selected to participate. Larger schools may have more than one class in a grade, in which case all classes, and thus students, in the selected grade participate. Not all teachers of a grade necessarily participate. Selection of teachers is decided within schools and is likely not random. When analyzing student and teacher outcomes we always consider all students and all teachers of the selected grade.

The randomization resulted in 130 eligible schools in the treatment group and 169 in the control group (Panel A in Table 1). All 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 Table

[15]As a result none of schools from the first 2015 cohort were considered. This legislation has since slightly changed to allow for directors to stay on longer.

[16]Other editions of the program included 4^{th} grade.

[17]*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.

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

1), 126 in the treatment group and 154 in the control group. The geographical distribution and treatment assignment of these schools is shown in Figure A4. Across the selected grades in each school, a total of 19,899 students were included in the experiment, 9,432 in treated schools and 10,467 in control schools (Panel C in Table 1).

Table 1: 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

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 data from the SEE, and collect data on cognitive and socio-emotional skills.

Administrative Data

We use administrative data from both the annual national school census and the state's education monitoring system to obtain school, teacher and student characteristics and progression.[19] It contains data on overall school characteristics, such as location, presence of a library, science

[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.

lab and internet, 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] The state's monitoring system, the *Sistema Integrado de Gestão da Educação* (SIGEduc) portal, provides data on passing, dropout, and retention rates at the grade level.[22] Where possible the analysis of the results uses both sources. Finally, the SEE provided data on school directors and on the implementation of the PIP, such as the score of the proposal, resources allocated to schools and execution of the project. Rate of implementation of the proposed plan is assessed by the mentor at each visit.

Learning Outcomes

To measure student learning, we use the standardized state exam in math, Portuguese, human and natural sciences, which were administered to 5^{th}, 6^{th}, 10^{th} and 12^{th} grades at the end of the 2016 school year. The RN standardized exam was introduced in 2016 and expanded to include all the PIP priority grades. For math and Portuguese, we obtain the scores rescaled to the national standardized test (SAEB), which allows us to put the impact on student learning in the Brazilian wide context.

Socio-Emotional Skills

To analyze the impact on socio-emotional skills, we measure the Big Five personality traits (neuroticism, extroversion, conscientiousness, agreeableness, openness).[23] We use a self-reported test developed and adapted to younger students in Brazil by the *Instituto Ayrton Senna* (IAS). This test, and equivalent, are widely used in the literature to assess socio-emotional skills.[24,25]

[20]We extract school location and distance from the capital of the state, Natal, by scraping 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 is collected and the survey of school, teacher and students' characteristics. In February-March of the following year, data is collected on passing/retention and on "movement", which includes dropout and transfers.

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

[23]The taxonomy of the five-factor model of personality we follow in this paper has been developed in the psychology literature following seminal work by Fiske (1949).

[24]See Kautz et al. (2014) for a review of the recent advances on 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 which help reveal the respondent latent scale and response style allowing us to calibrate the individual responses following the method suggested in Primi et al. (2016). The vignettes describe three hypothetical individuals that represent three clearly 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.

The test was administered at the end of the 2016 school year to the grade that entered the randomization (highest grade offered among 5th, 6th and 10th grade, see Section 3.1). In case a school had multiple classes in the same grade, one class was randomly chosen.

3.3 Validity of the Experiment

Randomization

To examine whether the randomization resulted in balanced samples across control and treatment groups, we compare observable characteristics prior to roll-out of the project. Table 2 shows several characteristics at the school, grade, teacher and student level, 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.[26] Generally we find no statistical differences when comparing the treatment and control groups. A joint significance test of school and student characteristics confirm 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. We find a statistically significant, yet small, difference in age of 6th graders. The control group is on average 0.25 years older than the treatment group. In the analysis we check robustness of the results to the inclusion of this unbalanced variable as a control.

Implementation

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, 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

[26] See Young (2019) on the importance of randomization statistical inference in experimental setups and Heß (2017) for a guideline on its implementation.

Table 2: Balance Table

	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	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 turnover rate	130	0.300	109	0.279	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 Família*	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]

Notes: 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 Gestão da Educação* (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 project, 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.

approval, all schools received the first visit of the mentor 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 along 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 balance with federal, state or municipal tax collection agencies.[27] The lack of this documentation delayed the transfer of funds for most schools. Transfers were supposed to take place at the beginning of the school year in February, but the first transfers were only made in July. By the end of the school year of 2016, 90 schools had received the funding.[28] Despite the challenges with the transfer of resources, mentors worked with the schools to continue implementation of the activities proposed in their work plan. By the end of the school year 79.37% of schools implemented at least 70% of the planned activities. All analysis take consideration the original assignment in the experiment (Panel B in Table 1) and should therefore be interpreted as intent-to-treat (ITT) effects.

Missing Data

Not all schools and students participated in the socio-emotional and proficiency test. Table A2 compares participation rates between the control and treatment group. Overall, 84% of schools participated in the socio-emotional test and 94% participated in the state standardized tests. Among the participating schools, on average, 55% of enrolled students took the socio-emotional test, and 69% participated in the proficiency tests.[29] Treated schools are more likely to participate in the socio-emotional test (91% versus 78%), 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% 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

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

[28] 8 schools received the funding in the following year.

[29] 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.

our results, we replicate the balance table restricting the sample to schools with at least one test-taker (Tables A3 and A4). We find similar balance results between treatment and control schools among this sub-sample of test-takers. To test whether school quality varies across test-takers and non-test-takers, we compare schools that participated in the socio-emotional test with those who did not, across treatment and control groups (Figure A5), using the 2015 IDEB as a measure of school quality at baseline. As expected, treatment and control schools generally have similar IDEB scores ($p = 0.98$). However, participating schools generally have better scores than non-participating schools ($p = 0.02$). Yet this pattern appears to be no different among treatment and control groups ($p = 0.62$). This suggest that our results are likely unbiased estimates of program impacts among tested schools, yet they may not extend to the non-tested schools.

4 Empirical Strategy and Results

4.1 Empirical Strategy

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

$$y_{is} = \alpha + \beta \cdot T_s + \Sigma_{strata} + \varepsilon_{is} \tag{1}$$

where y is the outcome of interest for student i in school s, T_s is the indicator variable of treatment assignment, Σ is a vector of strata dummies, and ε is the error term. Standard errors are clustered at the school level.[30] Since not all assigned schools received all components of the project, as discussed in Section 3.1, the parameter β measures the ITT effect. We provide estimates of project impact for all schools pooled as well as for each grade separately. To check robustness of the results, we estimate the model adding controls, and we use interaction-weighted, regression-weighted and blocked difference-in-means estimators.[31]

To explore potential distributional effects of the project we estimate unconditional quantile treatment effects (UQTE) effects following Firpo et al. (2009). Unlike the average effect, quantile

[30] Some estimates are obtained at the school level. In these cases, we employ robust standard errors.

[31] 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. The blocked difference-in-means approach uses strata sizes, instead of fixed effects, to weight the treatment effects estimates within each strata.

treatment effects assess whether the impact of the project differs at distinct points (quantiles) of the outcome distribution. The UQTE has a similar interpretation as the average effect and it is estimated by computing the horizontal difference between accumulated (or marginal) distributions of treated and control outcomes for a given quantile. For example, the effect on the median is given by $UQTE_{0.5} = Q_{0.5}(Y_T) - Q_{0.5}(Y_C)$, where Y_T is the value of the outcome variable (e.g., test score) at the distribution median of the treated group and Y_C is the value of the outcome variable at the distribution median of the control group.

4.2 Results

We begin the analysis with the key outcomes targeted by the project: student learning and progression indicators such as grade passing, repetition and dropout. In Section 5 we look at intermediate outcomes, such as teacher-turnover and socio-emotional skills to shed light on the potential mechanisms leading to impacts on final outcomes.

Learning Outcomes

Table 3 shows ITT estimates on overall test scores as well as separated by subject and grade. We find large positive impact on learning outcomes among schools assigned to treatment, but for 6$^{\text{th}}$ graders only. The intervention improved overall test scores for 6$^{\text{th}}$ graders by 0.15 SDs, or 6 points compared to a mean in the control group of 163. Significant improvements are observed across all subjects but are more pronounced for math and Portuguese. For robustness, we re-estimate the model controlling for a vector of covariates[32] and using alternative estimation strategies, such as interaction-weighted and regression-weighted estimators, blocked difference-in-means, as well as collapsing data at the school level. The results are very similar and are available in the Online Appendix.[33] The quantile regression estimates for 6$^{\text{th}}$ graders are presented in Figure A6 (average test score), and suggest gains across the board, with a more pronounced difference at the higher end of the grade distribution.[34]

[32]They include student's age, gender and race dummies (white, indigenous, black, or *pardo*), whether they receive *Bolsa Familia*, and whether they use school transportation.

[33]The Online Appendix can be accessed through this link.

[34]Quantile results disaggregated by subject are available in the Online Appendix.

Table 3: 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)
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)
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)
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

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. 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. 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.

In Figure A7, we compare both average and quantile treatment effects by gender. On average, PIP positively affected learning outcomes of both female and male 6th graders; however, distributional analysis suggests that the project shifted the entire distribution of test scores for boys to the right, but for girls resulted only in differences in the higher quantiles. We find suggestive evidence that the project 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 project in 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 constructed to allow for comparison of levels on a unique proficiency scale across grades and years.[35] This enables calculation of the accumulated knowledge in math and Portuguese of an average student between the tests taken in 5th and 9th grade. To calculate the average gains in knowledge between those 4 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 gains in test score for this cohort was 60 points, or 15 points per year on average. Based on the ITT estimates, we find that PIP improved 6th graders math and Portuguese scores by 6.81 points on the SAEB exams scale, the equivalent of a little under half a year of additional schooling.[36] In Section 6 we reflect on the economic implications of these results.

Student Progression

Table 4 shows the effect on passing, retention and dropout rates across grades. We report results using both data from SIGEduc, which is reported at the grade level, and from tracking individual students using the 2016 and 2017 waves of the school census.

We find positive impacts on overall progression. These are driven by substantial improvements in 6th grade, which is consistent with the results on learning gains. Passing rates in 6th grade are estimated to increase by 8.46 pp, a 13% improvement compared to the control mean of 63.56%. We find similar results using the census data; a 7 pp increase among 6th graders (12%). We

[35] 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.

[36] The OLS results in terms of SDs, using SAEB-rescaled test scores as outcome variable, are available in the Online Appendix. In our data, one SD deviation 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.

find no evidence of differential impact by gender (Table A5) or by baseline levels of passing rate (Table A6). The latter estimates confirm that the provision of schools' relative performance during the design workshops likely did not drive the results.

The impacts on grade passing mechanically result from either a reduction in dropout or retention or a combination of both. The SIGEduc data suggests that the result was mainly achieved by reducing grade repetition by 6.85 pp (23%). However, census data point to a reduction in dropout being the main driver. The discrepancy in the results can be explained by the difference in timing of defining a student's status. The SIGEduc data only captures students dropping out during the school year, while the census also captures dropout of students over the summer break. This suggest that part of the students reported as retained in SIGEduc drop out by the beginning of the next school year.

Table 4: 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)
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)
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
Retention								
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)
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

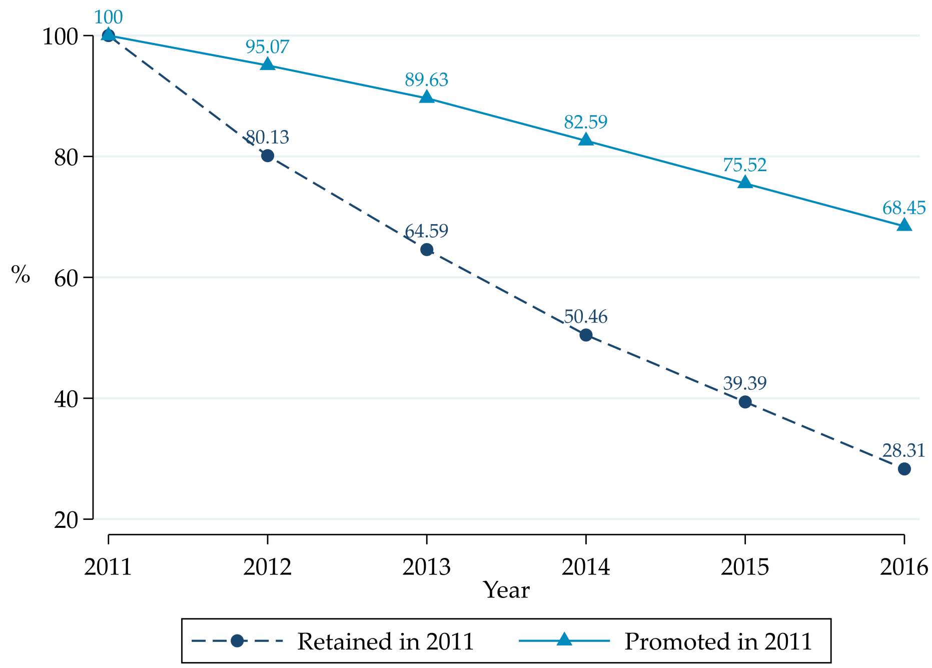
Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. School-level data are from *Sistema Integrado de Gestão da Educação* (SIGEduc) and student-level data are from Rio Grande do Norte census. Unit of observation: school and student. Outcome variables in the column headers. All regressions 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.

The reduction in 6$^{\text{th}}$ grade retention might have long-term implications for students' years of education and likelihood of completing school. To evaluate how much improving progression may affect students' school careers we track all RN students that were in 6$^{\text{th}}$ grade in 2011 up to 2017 using school census data. We find that students who were promoted in 6$^{\text{th}}$ grade in 2011 are 40 pp more likely to be in school in 2017 than students who were retained in 2011 (Figure 1a). Similarly, after 6 years, they have completed 2.34 more years of schooling (Figure 1b). We quantify the correlation between retention in 6$^{\text{th}}$ grade and schooling outcomes by estimating a simple OLS regression of dropout and completed years of schooling on retention.[37] We find that failing 6$^{\text{th}}$ grade is associated with a 21 pp higher likelihood of school dropout after 6 years, and a reduction of 1.7 years of completed schooling (Table A7). Taken at face value, our estimates provide suggestive evidence that the reduction of 23% (or 7 pp) in repetition rate caused by the PIP might contribute to substantially reduce school dropout (by 4.83 pp) and increase years of schooling (by 0.4 extra years) of the treated cohort of 6$^{\text{th}}$ graders.

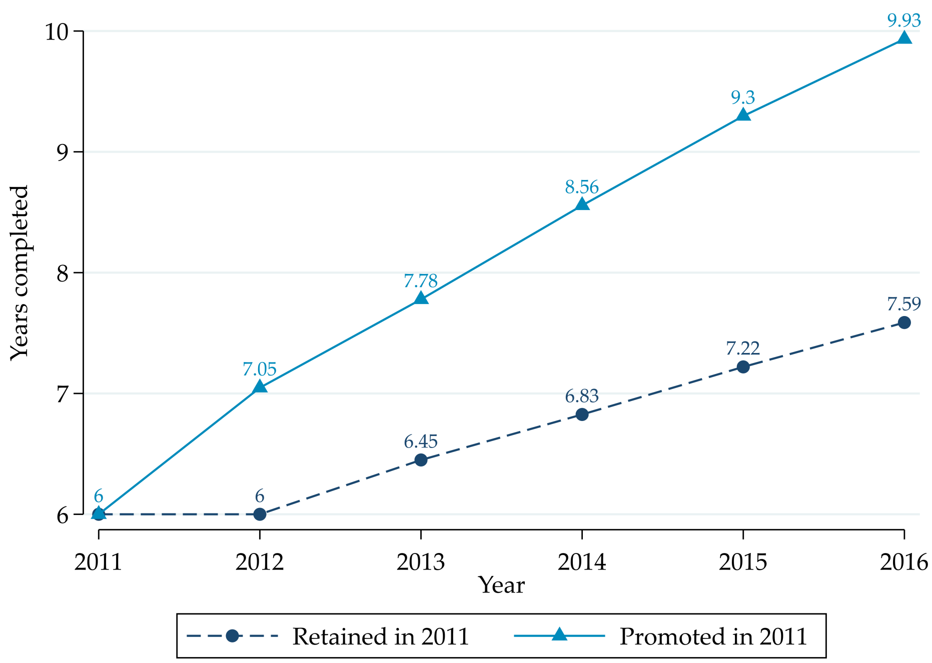
[37]We estimate the following cross-section regression: $y_{isc} = \alpha + \beta \cdot retained_{isc} + \sigma_s + \gamma_c + \varepsilon_{isc}$, where y_{isc} is the outcome variable, i.e., dropout dummy or years of completed schooling, of student i in school s and class c, $retained_{isc}$ is a dummy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

Figure 1: 6^{th} Grade Retention and Student Attainment

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



(b) Years of Completed Schooling of 2011 6^{th} Graders



Notes: The points in Panel (a) show the percentage of 6^{th} graders in 2011 who were enrolled in any grade (6^{th} or higher) in the following years, up to 2016. Panel (b) shows the average years of completed schooling of students who were enrolled in 6^{th} grade in 2011 by each following year, up to 2016. The sample is the universe of students of 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.

5 Potential Mechanisms

The results show that PIP had substantial impact on student learning and progression in Grade 6. In this section we explore the mechanisms through which the intervention may have affected these outcomes. As discussed in Section 2.2, there are three main components to the project, namely: i) teacher autonomy, ii) technical support through workshops and mentors, and iii) financial support. We empirically investigate two, potentially complementary, mechanisms through which these can improve student outcomes:

1. Reducing teacher turnover by increasing teacher motivation through the provision of autonomy and resources to develop their own project with technical support;
2. Building student socio-emotional skills either directly through i) the implementation of innovative projects in the schools, which aim to enhance student-teacher interactions, increasing students' motivation and skills; or indirectly through ii) the impact on cognitive skills, or iii) as a result of the first mechanism.

Teacher autonomy over the design and use of the grant likely affects student outcomes through both mechanisms: 1) it may crowd in teacher intrinsic motivation and affect teaching quality as well as 2) lead to better locally tailored projects. The second mechanisms assumes that innovative pedagogical projects, aimed at changing student-teacher interactions, could generate positive results, regardless of teacher autonomy.

In this section we explore which of the two channels were likely affected. However, we cannot tell apart the relative importance of the three different project components or the two mechanisms, as the same package was offered to all treatment schools.

5.1 Teacher Turnover

PIP allowed substantial autonomy for teachers, which may have affected teachers' motivation and engagement with students. We do not directly observe teachers' motivation in our data. To document whether the PIP might have affected teachers' commitment and motivation, we look at teacher turnover as a proxy.[38]

[38] To define 'teacher turnover', we track teachers across years in the school census. The outcome is a dummy of whether a teacher is in the same school in two consecutive years. The dummy is zero if a teacher is still teaching in the same school (in any grade) and one otherwise.

ITT estimates presented in Panel A of Table 5 suggest the project increased the probability of a teacher staying in the same school the following year by 6.4 pp (i.e., a 20.7% decrease in teacher turnover over the control mean of 30.9%) in 6^{th} grade. The higher teacher turnover in control schools is driven by more teachers leaving to other schools, they do not leave the education system more than in treatment schools. We interpret this result as suggestive evidence that part of project's success in Grade 6 was achieved by increasing motivation among teachers.

To explore whether affecting teacher turnover is driving impact on final outcomes, we estimate heterogeneous effects by teacher turnover at baseline. Panel B in Table 5 suggests that the reduction in teacher turnover is concentrated in schools with high teacher turnover rates at baseline. 'High teacher turnover' is defined at the grade level and equal to one if that school has a turnover rate above the sample median of that grade before the intervention.[39] The median turnover at baseline in 6^{th} grade is 33.33%.[40] We leverage this finding to document whether teacher turnover is a likely mechanism driving the learning results. We indeed find that the impacts on learning and dropout are also concentrated in schools with high teacher turnover at baseline. Impacts on learning for this group approaches 0.28 SD (Table 6).

[39] To define 'high teacher turnover' schools at baseline we calculate the proportion of teachers in a grade who leave a school between the 2015 and 2016 school year. 'High teacher turnover' is defined as a dummy, which takes the value one if the proportion of teachers leaving that school is above the median turnover distribution for schools treated in that grade.

[40] Most 5^{th} grade schools have only one teacher and their median turnover rate is zero. Therefore, we are not able to estimate heterogeneous effects for this grade.

Table 5: Impact on Probability of Teacher Staying in the Same School

	(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)
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 turnover at baseline				
Treatment	-0.016		0.014	-0.038
	(0.042)		(0.046)	(0.064)
Treatment × High teacher turnover rate at baseline	0.095*		0.104	0.139
	(0.058)		(0.074)	(0.092)
High teacher turnover rate at baseline	-0.109***		-0.127**	-0.115*
	(0.038)		(0.050)	(0.058)
Constant	0.776***		0.756***	0.780***
	(0.025)		(0.032)	(0.036)
Total effect on schools with high turnover at baseline: Treatment + Treatment × high-turnover dummy				
$\sum \hat{\beta}$	0.079		0.118	0.101
P-value	0.048		0.043	0.131
Unconditional mean of the dependent variable in the control group:				
Schools with high turnover at baseline	0.664		0.626	0.649
Schools with low turnover at baseline	0.775		0.762	0.786

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. Data are from Rio Grande do Norte 2016 and 2017 teacher censuses. Unit of observation: teacher. Outcome variables in the column headers. $\sum \hat{\beta}$ 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. Note that the coefficient on the high-turnover dummy at baseline for 5^{th} grade is not identified because the median itself is equal to 0. This is due to the fact that, in most schools, 5^{th} grade has only one teacher, thus the school turnover rate variable is either equal to 0 or 1.

Table 6: Impact on Student Learning and Progression by Teacher Turnover at Baseline

	Learning	*Progression*		
	(1) Average test score	(2) Passed	(3) Dropped out	(4) Retained
All schools				
Treatment	-0.059	0.081**	0.006	-0.086***
	(0.075)	(0.037)	(0.025)	(0.029)
Treatment × High teacher turnover at baseline	0.132	-0.063	-0.023	0.087**
	(0.093)	(0.046)	(0.031)	(0.035)
High teacher turnover at baseline	-0.188***	0.046	0.016	-0.062**
	(0.072)	(0.034)	(0.026)	(0.028)
Constant	0.118*	0.569***	0.162***	0.269***
	(0.066)	(0.025)	(0.020)	(0.023)
Number of observations	11794	16159	16169	16159
Number of clusters	228	239	239	239
Total effect: Treatment + Treatment × High teacher turnover at baseline				
$\sum \hat{\beta}$	0.073	0.017	-0.018	0.000
P-value	0.176	0.541	0.328	0.982
6th grade – Lower secondary schools				
Treatment	0.016	0.079*	-0.025	-0.054
	(0.080)	(0.043)	(0.027)	(0.037)
Treatment × High teacher turnover at baseline	0.261**	-0.023	-0.043	0.066
	(0.119)	(0.058)	(0.037)	(0.053)
High teacher turnover 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 × High teacher turnover at baseline				
$\sum \hat{\beta}$	0.276	0.055	-0.067	0.012
P-value	0.002	0.186	0.008	0.752

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.

To assess whether solely reducing teacher turnover in itself is sufficient to achieve these results, we explore the fact that many 6th grade teachers also teach in other grades, where no innovative projects are implemented. According to the school census data, 90.43% of 6th grade teachers also teach 7th grade, 81.76% in 8th grade, and 73.21% in 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 of Table 7). We compare student level outcomes for 6th grade schools, in their remaining lower-secondary grades (Panel B of Table 7).[42] We only have access to data on student progression in other grades, the standardized test was not implemented in 7th grade. The lack of positive impacts on other grades suggests that reducing teacher turnover alone might not be sufficient to affect student outcomes: positive results in 6th grade are likely driven by the combination of increased motivation of teachers and the other project components.[43] Moreover, 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.

[41] The percentage is balanced between treatment and control schools.
[42] The results using grade level data from SIGEduc are very similar. See Online Appendix.
[43] Results by teacher turnover at baseline also show no impacts on other grades. See Online Appendix.

Table 7: Impact on Other Grades in 6^{th} 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 being retained			
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

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. 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 6^{th} 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.

5.2 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 the traditional lecture-based teaching. As a consequence, resulting projects may have directly affected student socio-emotional skills. In addition, increasing teacher motivation and commitment may provide an indirect channel to improve socio-emotional skills.

Table 8 shows the ITT estimates on each of the Big Five personality traits. The indicators are standardized (within grade) and the coefficients can be interpreted in terms of standard deviations. Pooling all grades, we find that the project had a positive and statistically significant effect on conscientiousness and extroversion.[44] However, in line with previous results, these are driven by the impacts on 6$^{\text{th}}$ graders (0.17 SD and 0.21 SD respectively). Among the Big Five, the trait of 'conscientiousness' is commonly associated with acquisition of cognitive skills (Poropat, 2009; Ivcevic and Brackett, 2014). It encompasses traits such as self-control, organization, responsibility and perseverance.

We observe that student test scores and socio-emotional outcomes are positively correlated in the tested sample at endline, regardless of treatment status (Figure A8). Unfortunately, as mentioned in Section 3.2, data on socio-emotional skills were only collected for a random subset of students in each school,[45] therefore we cannot further investigate the mediating role of socio-emotional skills on learning outcomes or vice versa.[46] However, in line with the literature, these correlations confirm the complementarities between socio-emotional and cognitive skills.

5.3 Understanding Heterogeneous Impacts by Grade

We presented the results to the mentors in a focus group discussion to shed light on what may be driving differences in impacts across grades. First, we found that mentors had more experience with teaching and implementing projects in lower grades, which may have resulted in

[44] The same robustness checks used for estimating the impact on learning outcomes can be seen in the Online Appendix.

[45] Participation rate in the socio-emotional test and other observed characteristics are balanced at the student level when we restrict the sample to the subset of students who took the socio-emotional test. See discussion in Section 3.3.

[46] When we restrict the sample to the subset of students who took the socio-emotional test, we are unable to detect effects of the project on learning outcomes.

Table 8: 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)
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)
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)
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

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. 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. 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.

the technical assistance being better tailored to these grades. Second, mentors stated that the project filled a clear gap faced by 6$^{\text{th}}$ graders who experience a significant transition between levels of education. The key difference between these levels is that students in primary education 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 5). The potential negative impact of this transition is well documented in the US and has been recently investigated in Brazil (Bedard and Do, 2005; Cook et al., 2006; Hanewald, 2013; Santos et al., 2017). Santos et al. (2017) evaluate the impact of a pilot in municipal schools in Rio de Janeiro, Brazil, which expanded primary school to include 6$^{\text{th}}$ grade. They find that having the 6$^{\text{th}}$ grade in the primary school increases learning by 0.16 SD, and suggestive evidence that strengthening of student-teacher relationship mediated some of the effect on learning.

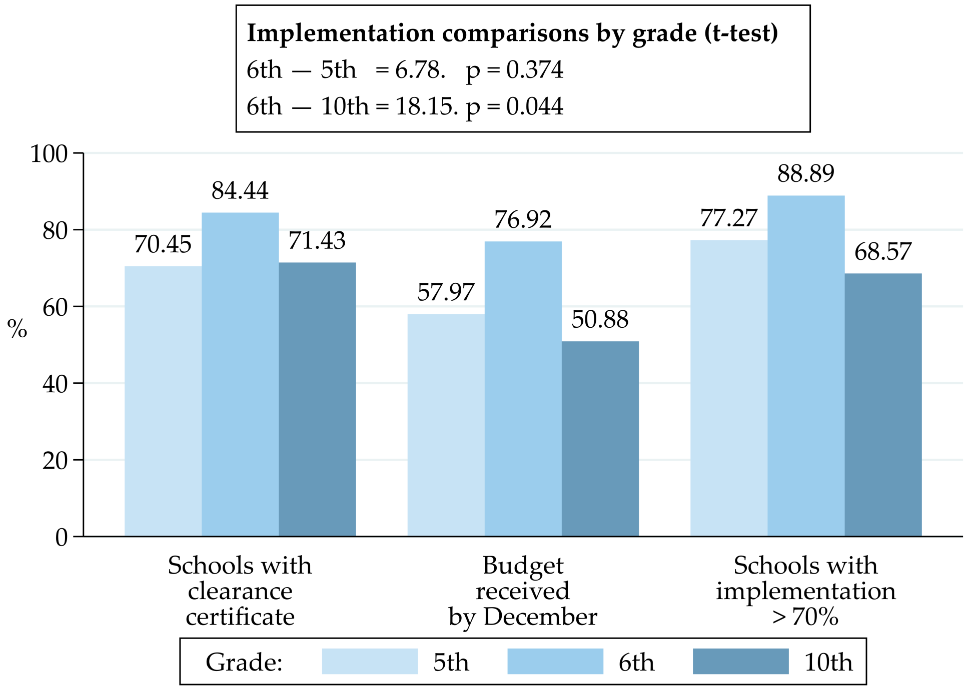
We compare administrative data to assess whether project implementation varied across the three grades. All treatment schools received similar levels of support from the SEE team: all had an approved sub-project and were assigned a mentor who visited them regularly. Here we focus on the implementation of the planned activities by the schools throughout the year. We report three measures of implementation: i) obtaining the clearance certificate, which is a necessary requirement for schools to receive funding from any state level educational program[47]; ii) percentage of project funds received by the end of the school year; and iii) whether a school implemented at least 70% of the planned activities included in the work plan. We observe substantial difference in rates of implementation across the three grades (Figure 2). Each of the indicators shows higher rates of implementation in 6$^{\text{th}}$ grade.[48]

Taken together, Grade 6 schools may have perceived the project as being particularly relevant to smooth shocks observed around the transition from primary to lower secondary education. This may have led to better implementation in this grade.

[47] We indeed find that obtaining the clearance certificate is what most predicts rate of implementation (Table A8). We find that being assigned to receive the PIP increases the likelihood of schools obtaining the clearance certificate by 41 pp. This impact does not differ by grade (Table A9).

[48] We do not observe any significant correlation between school characteristics at baseline and implementation (Table A8), however it is likely that implementation is endogenous to unobserved school quality and our outcomes of interest, therefore we refrain from comparing schools with high rates of implementation with low rates of implementation as this would provide biased results.

Figure 2: Implementation by Grade



Notes: '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. Data are from State Secretariat of Education (SEE). Sample: schools treated.

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, Figure A9 shows that if students retain their learning gains over time, as measured by SAEB scores, the impact of the PIP would suffice to 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 A10. The strategy is described in more detail in Appendix B.1.

Expected Returns to Education

We expect the intervention to impact labor market outcomes of the 6^{th} graders in the long term through two channels: first via learning gains among those that stayed in school (productivity channel), and second via 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-envelop calculation suggests a net present value (NPV) on future earnings of 29,148.97 Brazilian Reais (BRL) – or 7,287.24 US\$. 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 a NPV on future earnings of 23,319.18 BRL (or 5,829.79 US\$). The full effect on expected earnings would then range from 7 to 13 thousand US dollars or 28 to 52 Brazilian minimum wages. The data and methodology used for the calculations are described in Appendix B.2.

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

7 Conclusion

In this paper, we study whether providing autonomy to public sector agents, such as teachers, can improve the quality of service delivery in a low state capacity environment. The possibility of stimulating decision-making responsibility of local public officials to make best use of their contextual knowledge to design and implement more effective policies is a first order question in the public sector, especially when resources are scarce and the monitoring capacity of the central government limited.

We explore this question in the context of an education program, which was randomly rolled out in state schools of Rio Grande do Norte, one of the poorest Brazilian states. The Pedagogical

Innovative Program (PIP) consisted of three key components: i) teacher autonomy to develop a work plan to tackle problems locally identified; ii) technical assistance from mentors assigned by state secretariat to support teachers during the diagnostic and development of the work plan stages; and iii) funding earmarked for schools to buy the pedagogical material necessary to implement the activities described in the work plan. The project was designed to motivate teachers and students and improve students' outcomes, leveraging mostly existing staff and school resources.

We find that the PIP had meaningful impacts on 6$^{\text{th}}$ graders, a critical grade during the transition from primary to lower-secondary education. Our ITT estimates point to learning gains in math and Portuguese of 0.18 SD and 0.16 SD respectively. In addition, we find that passing rates increased whereas dropout and retention decreased. To shed light on the mechanisms underpinning our main results we tested whether the program affected teacher turnover and students' socio-emotional skills as the program envisaged teachers and students motivation as main pathways for program's success. The program reduced teacher turnover by 20.7% and that most of this reduction was observed in schools with higher teacher turnover at the baseline. Consistent with these results, we document learning gains almost twice as high in schools with high teacher turnover at the baseline. To estimate impacts on the socio-emotional skills, we use the Big Five personality traits. Our results show positive impacts on conscientiousness and extroversion. Overall, these findings empirically support the program's intention to impact students' outcomes by motivating teachers and students.

These results have direct implications for policy design in countries that might neither have fiscal space to design pay-for-performance schemes at scale nor effective monitoring mechanisms. Autonomy over limited funding appeared to be enough to provide a non-monetary incentive to increase teacher motivation. In combination with the technical assistance the program mitigated agency problems observed in other types of interventions where the decentralization of decision-making to local officials backfired (e.g., Banerjee et al., forthcoming) while complementing local capacity.

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 motivation of agents and final recipients, in this case students, is particularly important to affect outcomes. More research is needed to understand in which settings this approach is more likely to succeed.

References

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011): "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *The Quarterly Journal of Economics*, 126, 699–748. ↑5

Akhtari, M., D. Moreira, and L. Trucco (2018): "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services," Working Paper. ↑4, ↑6

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. ↑5

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education*, Elsevier, vol. 4, 1–181. ↑5

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. ↑7

Amabile, T. M., W. DeJong, and M. R. Lepper (1976): "Effects of Externally Imposed Deadlines on Subsequent Intrinsic Motivation." *Journal of Personality and Social Psychology*, 34, 92. ↑2

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. ↑2

Ashraf, N., O. Bandiera, and B. K. Jack (2014): "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery," *Journal of Public Economics*, 120, 1–17. ↑5

Bandiera, O., M. C. Best, A. Q. Khan, and A. Prat (2020): "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," NBER Working Paper No. 26733. ↑3

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (forthcoming): "Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training," *American Economic Journal: Economic Policy.* ↑2, ↑34

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. ↑5

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. ↑5, ↑30

Blimpo, M. P., M. Blimpo, D. Evans, and N. Lahire (2015): "Parental Human Capital and Effective School Management: Evidence from the Gambia," World Bank Policy Research Working Paper No. 7238, The World Bank: Washington, DC. ↑5

Bowles, S. and S. Polania-Reyes (2012): "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50, 368–425. ↑5

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. ↑2

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. ↑5

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. ↑2

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104, 2593–2632. ↑2

——— (2014b): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104, 2633–79. ↑2

Cook, P. J., R. MacCoun, C. Muschkin, and J. Vigdor (2006): "Should Sixth Grade be in Elementary or Middle School? An Analysis of Grade Configuration and Student Behavior," NBER Working Paper No. 12471. ↑30

Cunha, F. and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47. ↑4

Cunha, F. and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782. ↑4

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931. ↑4

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. ↑5

Deci, E. L. (1971): "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology*, 18, 105. ↑2

Deci, E. L. and R. M. Ryan (1985): *Intrinsic Motivation and Self Determination in Human Behaviour*, Plenum, New York. ↑2

Dobbie, W. and R. G. J. Fryer (2013): "Getting Beneath the Veil of Effective Schools: Evidence from New York City," *American Economic Journal: Applied Economics*, 5, 28–60. ↑5

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. ↑2

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 102, 1241–78. ↑5

Evans, D. and A. Popova (2016): "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews," *World Bank Research Observer*, 31, 242–270. ↑2

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. ↑19, ↑62

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973. ↑16

Fiske, D. W. (1949): "Consistency of the Factorial Structures of Personality Ratings from Different Sources," *The Journal of Abnormal and Social Psychology*, 44, 329. ↑12

Fryer, R. G. J. (2014): "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *The Quarterly Journal of Economics*, 129, 1355–1407. ↑5

——— (2017): "Management and Student Achievement: Evidence from a Randomized Field Experiment," NBER Working Paper No. 23437. ↑5

Gibbons, C. E., J. C. S. Serrato, and M. B. Urbancic (2018): "Broken or Fixed Effects?" *Journal of Econometric Methods*, 8. ↑16

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. ↑2

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. ↑5, ↑30

Hanushek, E. A., S. Link, and L. Woessmann (2013): "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA," *Journal of Development Economics*, 104, 212–232. ↑2

Hanushek, E. A. and L. Woessmann (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–68. ↑2

——— (2012): "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," *Journal of Economic Growth*, 17, 267–321. ↑2

Heckman, J. J. (2000): "Policies to Foster Human Capital," *Research in Economics*, 54, 3–56. ↑6

Heckman, J. J. and T. Kautz (2012): "Hard Evidence on Soft Skills," *Labour Economics*, 19, 451–464. ↑6

Heckman, J. J. and Y. Rubinstein (2001): "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *American Economic Review*, 91, 145–149. ↑5

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482. ↑5

Hess, S. (2017): "Randomization Inference with Stata: A Guide and Software," *The Stata Journal*, 17, 630–651. ↑13

Ivcevic, Z. and M. Brackett (2014): "Predicting School Success: Comparing Conscientiousness, Grit, and Emotion Regulation Ability," *Journal of Research in Personality*, 52, 29–36. ↑4, ↑29

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. ↑2

Jackson, C. K., J. E. Rockoff, and D. O. Staiger (2014): "Teacher Effects and Teacher-Related Policies," *Annual Review of Econonomics*, 6, 801–825. ↑4

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. ↑12

Lavy, V. (2009): "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *American Economic Review*, 99, 1979–2011. ↑5

Lindqvist, E. and R. Vestman (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3, 101–28. ↑5

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, 1627–1673. ↑5

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. ↑2

Mincer, J. (1974): *Schooling, Experience, and Earnings*, National Bureau of Economic Research: Cambridge, MA. ↑62

Muralidharan, K. and V. Sundararaman (2011): "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119, 39–77. ↑5

OECD (2015): "PISA: Programme for International Student Assessment," Accessed on 14 August 2019. ↑6

——— (2016): *Education at a Glance 2016*, OECD Publishing: Paris. ↑3, ↑8

Oliveira, L. F. B., R. Terra, et al. (2016): "Impacto do Programa Mais Educação em Indicadores Educacionais," . ↑5

Poropat, A. E. (2009): "A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance," *Psychological Bulletin*, 135, 322. ↑4, ↑29

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. ↑12

Pritchard, R. D., K. M. Campbell, and D. J. Campbell (1977): "Effects of Extrinsic Financial Rewards on Intrinsic Motivation," *Journal of Applied Psychology*, 62, 9. ↑2

Psacharopoulos, G. and H. A. Patrinos (2018): "Returns to Investment in Education: A Decennial Review of the Global Literature," Policy Research Working Paper; No. 8402, The World Bank: Washington, DC. ↑62

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. ↑2

Rasul, I., D. Rogger, and M. J. Williams (2018): "Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service," Policy Research Working Paper No. 8595, The World Bank: Washington, DC. ↑2

Rockoff, J. E., D. O. Staiger, T. J. Kane, and E. S. Taylor (2012): "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools," *American Economic Review*, 102, 3184–3213. ↑5

Rogger, D. and R. Somani (2018): "Hierarchy and Information," Policy Research Working Paper No. 8595, The World Bank: Washington, DC. ↑2

Ronfeldt, M., S. Loeb, and J. Wyckoff (2013): "How Teacher Turnover Harms Student Achievement," *American Educational Research Journal*, 50, 4–36. ↑4

Ryan, R. M. and E. L. Deci (2017): *Self-determination theory: Basic psychological needs in motivation, development, and wellness*, Guilford Publications, New York. ↑2

Sanchez Puerta, M. L., A. Valerio, and M. G. Bernal (2016): *Taking Stock of Programs to Develop Socioemotional skills: A Systematic Review of Program Evidence*, The World Bank: Washington, DC. ↑6

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. ↑3, ↑5, ↑30

Taylor, E. S. and J. H. Tyler (2012): "The Effect of Evaluation on Teacher Performance," *American Economic Review*, 102, 3628–51. ↑5

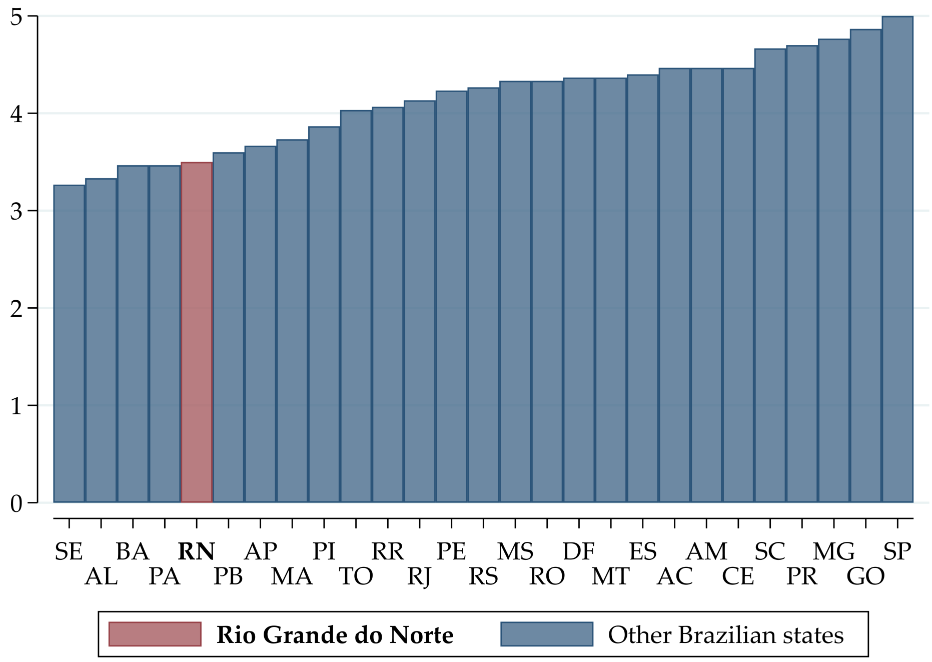
World Bank (2018): *World Development Report 2018: Learning to Realize Education's Promise*, The World Bank: Washington, DC. ↑2

Young, A. (2019): "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *The Quarterly Journal of Economics*, 134, 557–598. ↑13

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. ↑2

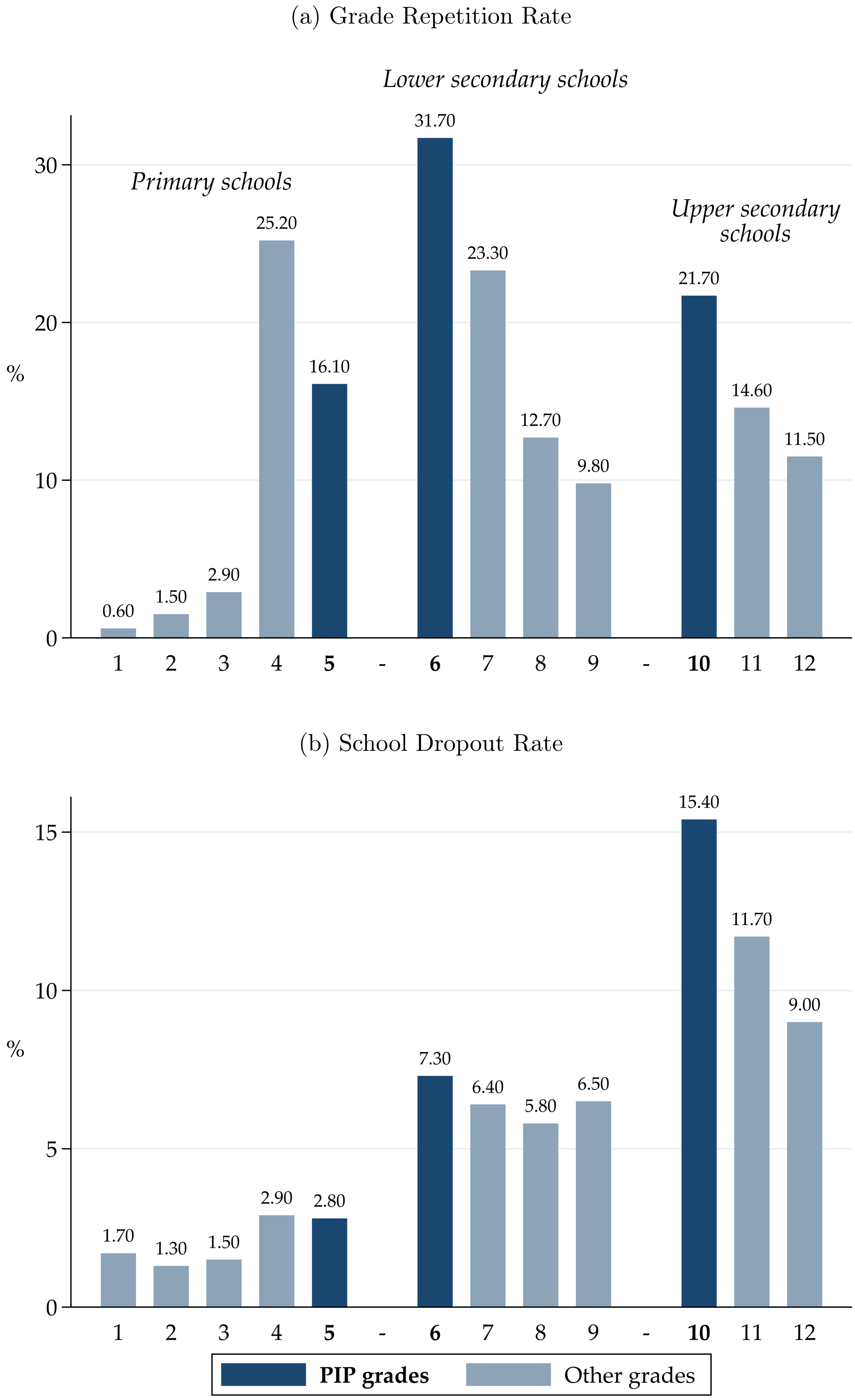
A Supplementary Figures and Tables

Figure A1: IDEB in Rio Grande do Norte vs. Other Brazilian States, 2015



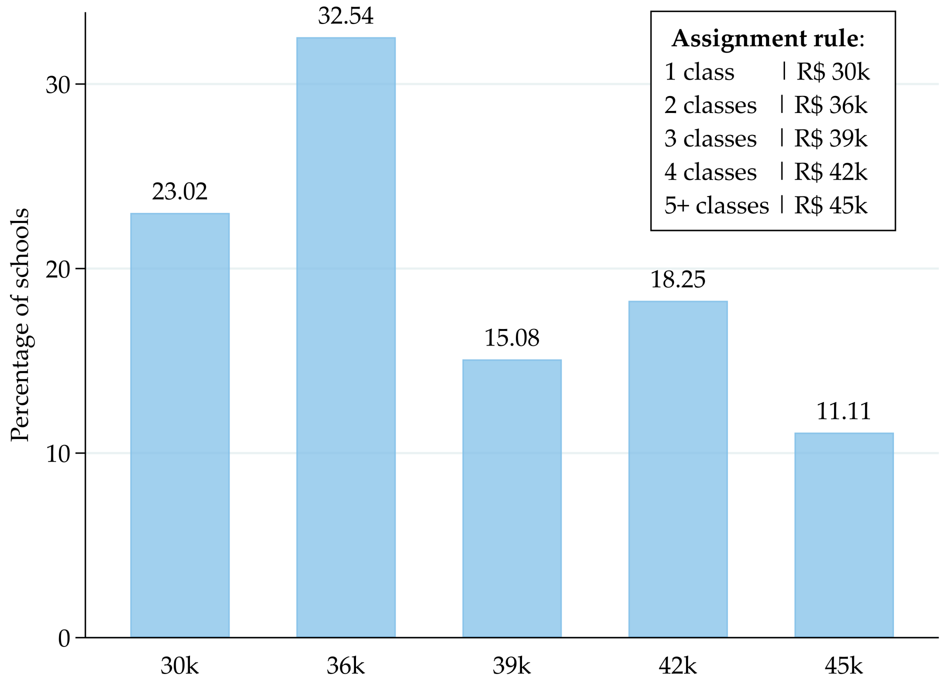
Notes: We use data from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) for state public schools. The IDEB index is defined at each education stage, i.e., for primary, middle, and secondary schools. It is a national indicator for the quality of education, which combines information on student test scores and passing rates (see Appendix B.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 A2: Grade Repetition and School Dropout Rates by Grade in Rio Grande do Norte



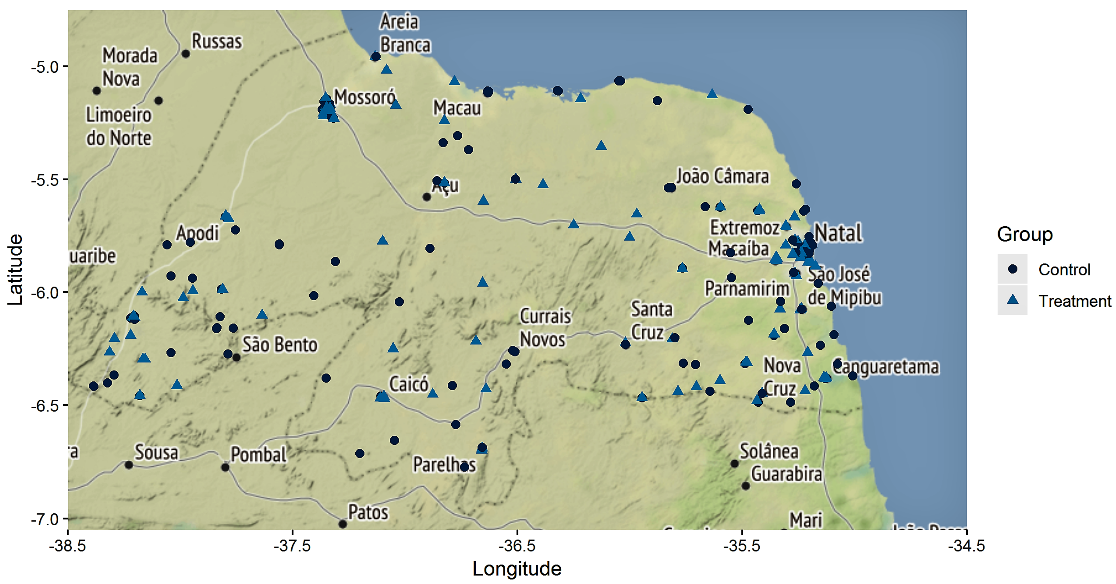
Notes: The bars show average retention and dropout rate among public schools in Rio Grande do Norte in 2015. Data are from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP).

Figure A3: Allocation of Resources by Type of Grant



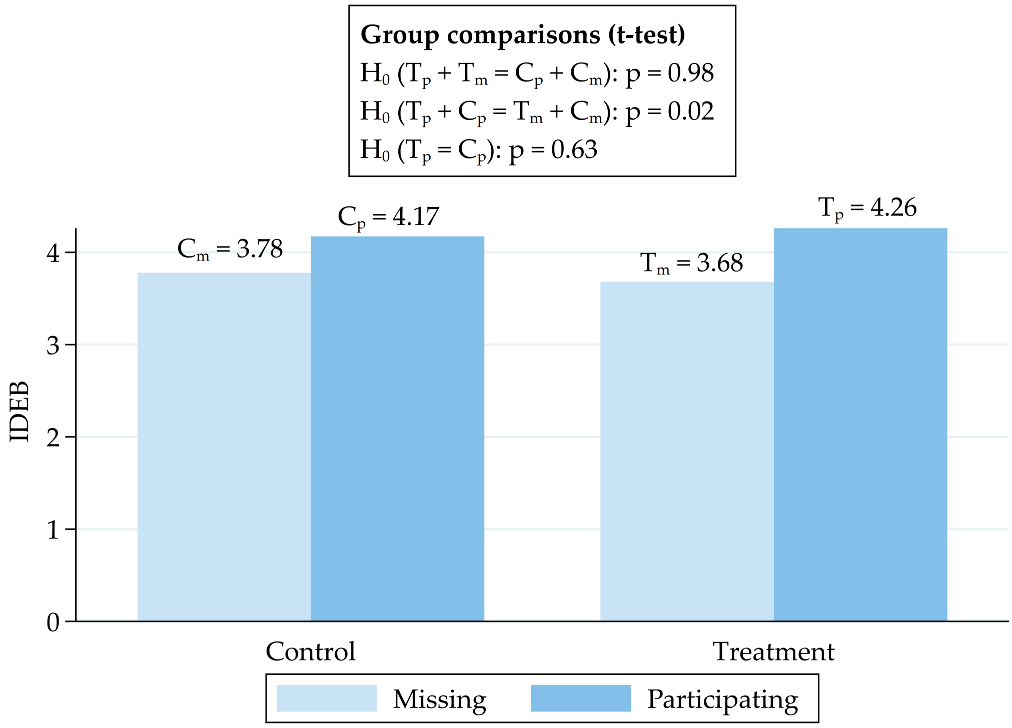
Notes: The bars show the percentage of schools by the type of grant they were assigned to receive through PIP (ranging from 30,000 to 45,000). The values are in Brazilian Reais, which were worth 0.25 US dollars at the beginning of 2016.

Figure A4: Geographical Distribution of Schools by Treatment Status



Notes: 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 A5: IDEB by Participation to Socio-Emotional Test and Treatment

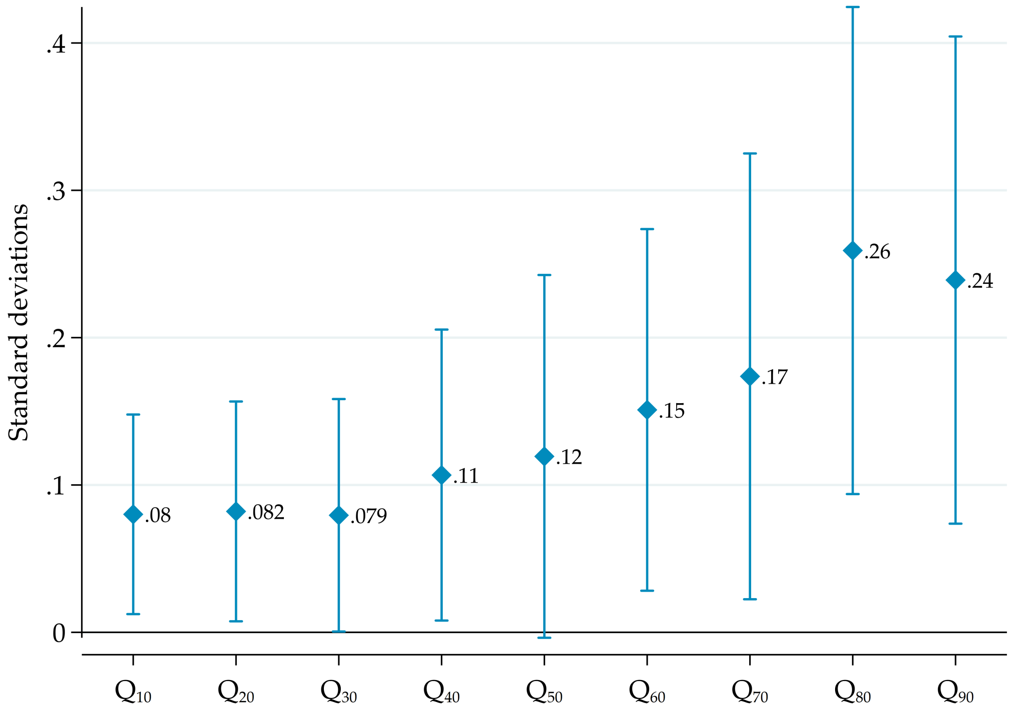


Notes: The bars show the unconditional means of the school IDEB by participation in the socio-emotional test and by treatment assignment, as described in 3.3. We regress IDEB on these 4 categories so that:

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

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).

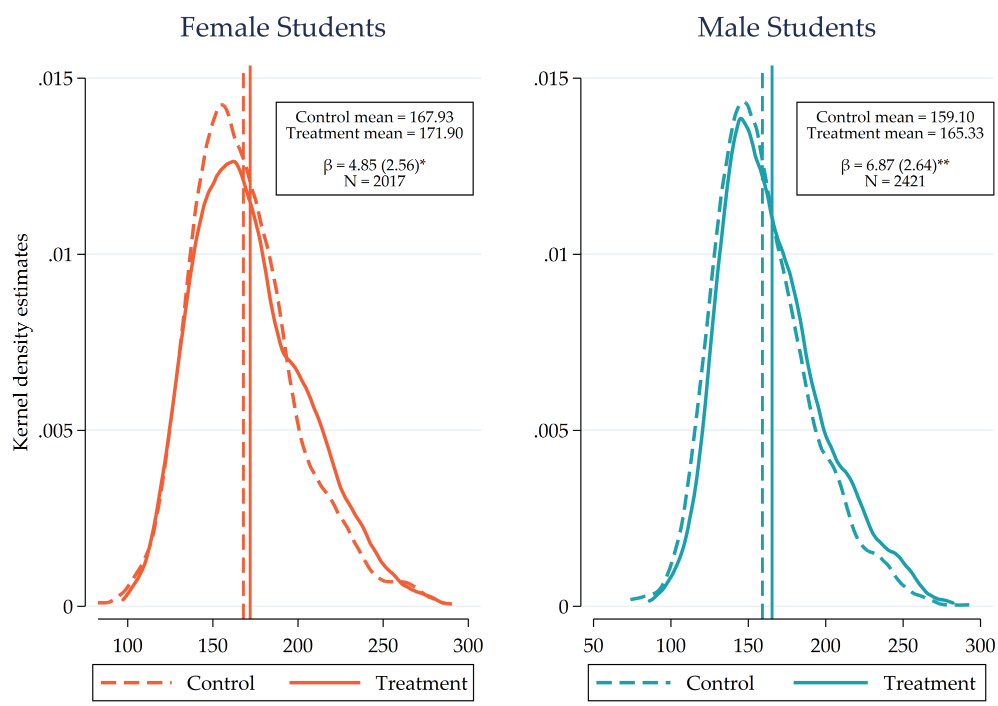
Figure A6: Quantile Treatment Effect on Average Test Score in 6^{th} Grade



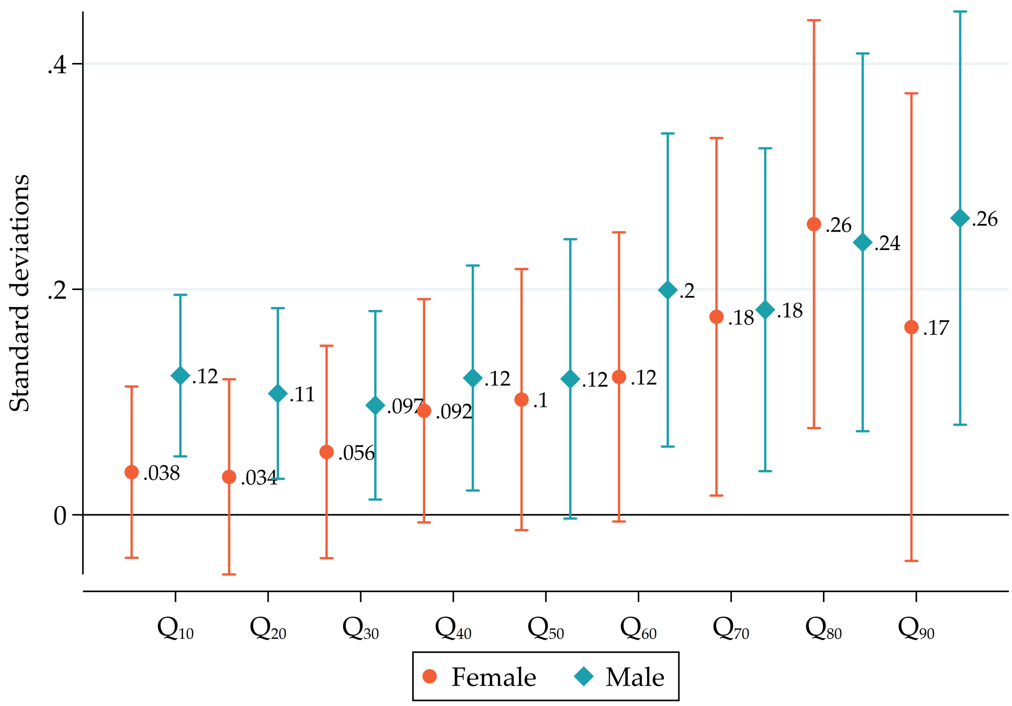
Notes: Point estimates of quantile regressions with strata (i.e., region) fixed effects and standard errors clustered at the school level. Confidence intervals are 90%. Sample: schools treated at 6^{th} grade.

Figure A7: Impact on Average Test Score by Gender in 6^{th} Grade

(a) Distribution



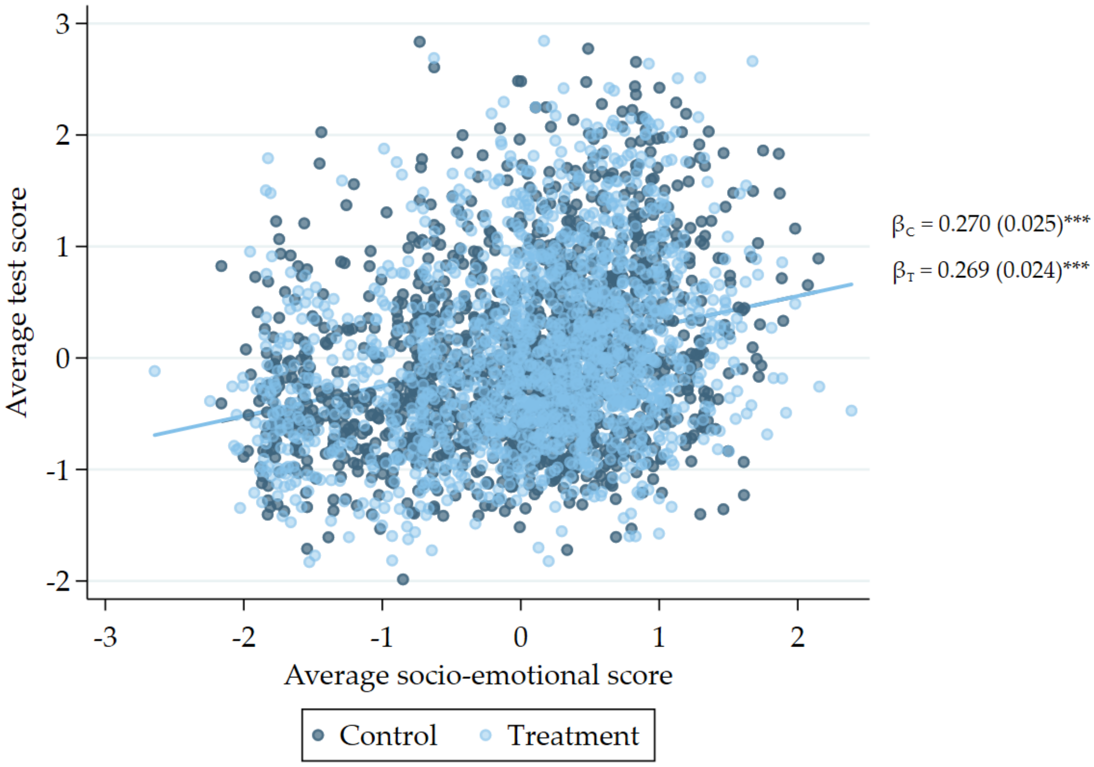
(b) Quantile Treatment Effect



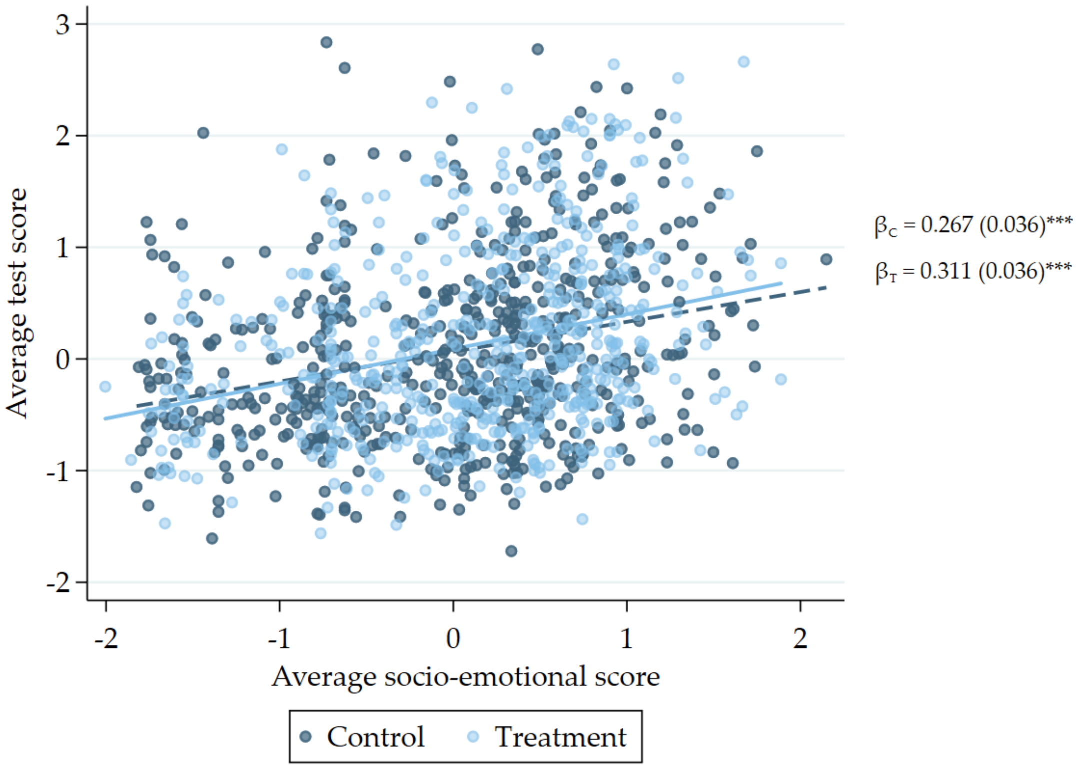
Notes: Average test score is the average of standardized test scores in math, Portuguese, human and natural science (range 0-400). Sample: schools treated at 6^{th} grade. Kernel densities are computed using Epanechnikov kernel function. Treatment effects in (a) are estimated through regressions with 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% confidence intervals. Quantile treatment effects are expressed in terms of standard deviations from the control group.

Figure A8: Scatter Plot of Cognitive and Socio-Emotional Skills

(a) All Schools



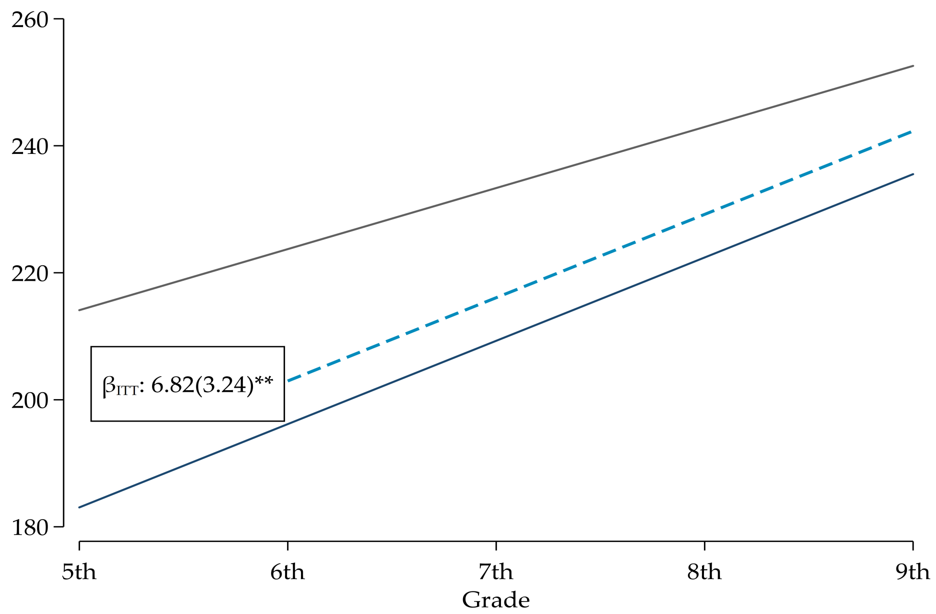
(b) 6th Grade



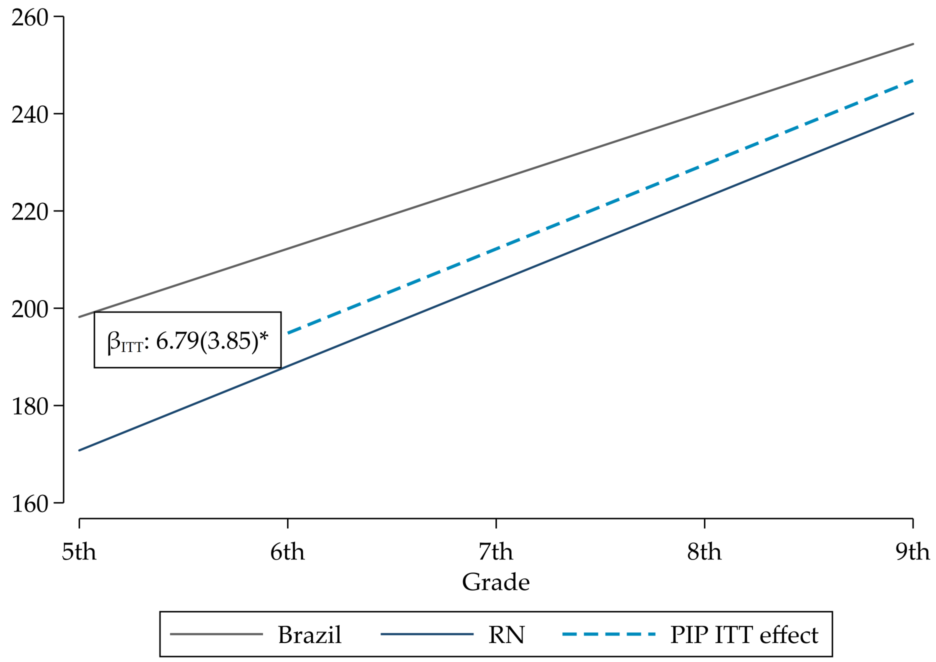
Notes: Unit of observation: student. The linear fits are estimated for both treatment and control group through an OLS regression with standard errors clustered at the school level. *** indicates significance at the 1 percent critical level. The sample is restricted to students who took the socio-emotional 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.

Figure A9: Learning Gains in 6^{th} Grade Rescaled to SAEB – Projection over Time

(a) Math

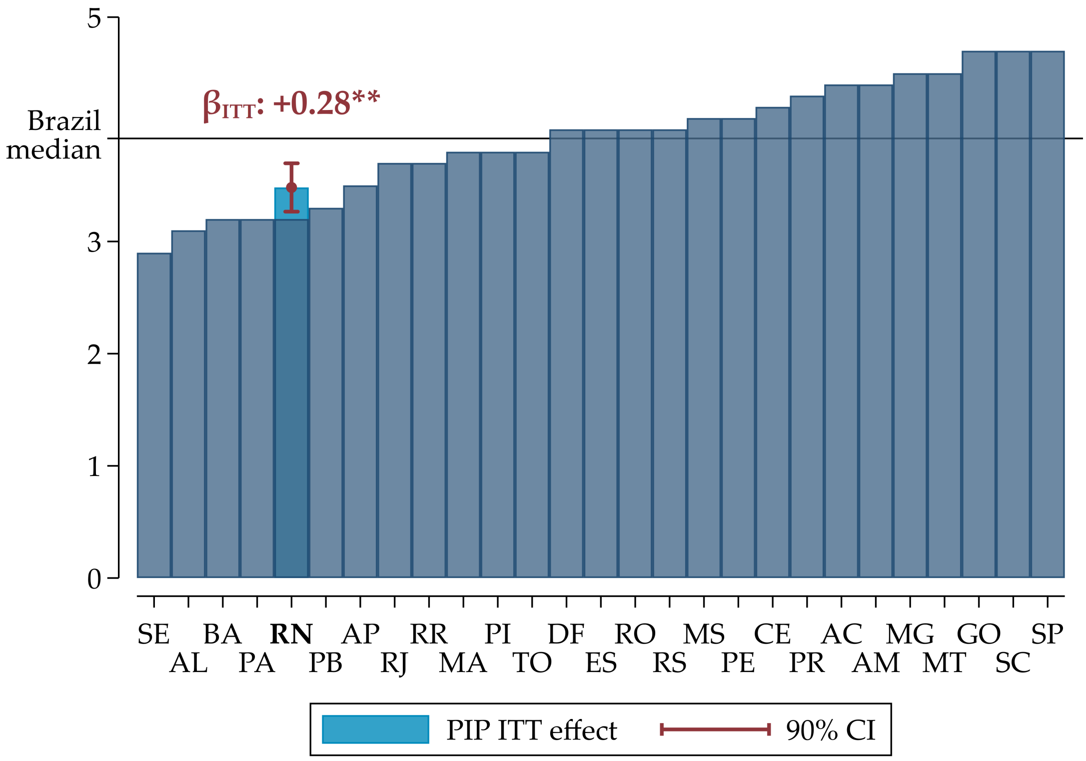


(b) Portuguese



Notes: 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 5^{th} grade in 2013 and 9^{th} grade in 2017. The points in 6^{th}, 7^{th}, and 8^{th} grades are linear interpolation. The PIP intent-to-treat effect on 6^{th} 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 A10: Learning Gains in 6^{th} Grade Rescaled to *IDEB* – Comparison with Other Brazilian States



Notes: We use data from *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) for state public schools. The bars show average IDEB by state in 2015. The PIP intent-to-treat effect on 6^{th} 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 Appendix B.1 for the methodology we follow to compute IDEB for our grades of interest.

Table A1: Effect of Teacher Permanence on Education Outcomes in Brazil

	(1) Age-grade distortion	(2) Passing rate	(3) Retention 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	✓	✓	✓	✓

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. 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 A2: 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)		

Notes: '*Participating schools*' is a dummy for schools which 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) 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 A3: Balance Table on Subsample of Schools with Socio-Emotional 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	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]
Retention 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 Família*	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]

Notes: 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 Gestão da Educação* (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 project, 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.

Table A4: 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]
Retention 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 Família*	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]

Notes: 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 Gestão da Educação* (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 project, 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.

Table A5: 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 being retained				
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

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. Data are from Rio Grande do Norte census. 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 A6: Impact on Student Progression Rates – Heterogeneity by Passing Rate at Baseline

	Grade level				*Student level*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 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-promotion 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-promotion 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
					Retention			
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-promotion 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

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. School-level data are from *Sistema Integrado de Gestão da Educação* (SIGEduc) and student-level data are from Rio Grande do Norte census. 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. Note that the coefficient on the high-turnover dummy at baseline for 5^{th} grade is not identified because the median itself is equal to 0. This is due to the fact that, in most schools, 5^{th} grade has only one teacher, thus the school turnover rate variable is either equal to 0 or 1.

Table A7: Impact of 6^{th} Grade Retention on Student Achievement

	Dropout			**Years of completed schooling**		
	(1)	(2)	(3)	(4)	(5)	(6)
Retention 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			✓			✓

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. 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 6^{th} grade students of 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. Data from 2011-2017 censuses.

Table A8: Drivers of Implementation

	(1)	(2)	(3)
Number of enrolled students in PIP grades	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)
Quality score of expression of interest	0.003	0.003	0.004
	(0.004)	(0.004)	(0.003)
School infrastucture index	-0.055	-0.049	-0.005
	(0.038)	(0.037)	(0.028)
Distance to Natal (km)	0.003**	0.003**	0.001
	(0.001)	(0.001)	(0.001)
Passing rate in 2015		-0.000	-0.000
		(0.002)	(0.002)
Dropout rate in 2015		0.004	0.002
		(0.006)	(0.003)
School has clearance certificate			0.617***
			(0.096)
Number of observations	123	122	122
Adjusted R-squared	0.161	0.149	0.517
Mean dep. var.	0.826	0.825	0.825
SD dep. var.	0.340	0.341	0.341

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. Data are from 2015 Rio Grande do Norte school census, State Secretary of Education and Culture (SEEC), and *Sistema Integrado de Gestão da Educação* (SIGEduc). 'Implementation' is defined as a school having the ratio of the number of activities that were implemented over the number of planned activities described in the work plan above 70%. 'School infrastructure index' is constructed through principal component analysis of the following dummy variables: whether school has internet, library, 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 in parentheses.

Table A9: 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	

Notes: *Significant at 10%. **Significant at 5%. ***Significant at 1%. Data are from 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.

B Back-of-the-Envelope Estimations

B.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.[49] 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, middle, and high schools, i.e., 5th, 9th, and 12th grades, are evaluated. SAEB tests are based on IRT so 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.[50]

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

$$N_{sj} = \frac{score_{sj} - min_j}{max_j - min_j} \cdot 10 \tag{B1}$$

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

$$min_j = \mu_j - 3 \cdot \sigma_j \,; \quad max_j = \mu_j + 3 \cdot \sigma_j \tag{B2}$$

[49] 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.

[50] 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.

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

$$N_s = \frac{N_{s,j=math} + figureN_{s,j=Portuguese}}{2} \tag{B3}$$

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,

$$T_s = \frac{\sum_{y=1}^{Y} \frac{1}{p_{sy}}}{Y} \tag{B4}$$

$$P_s = \frac{1}{T_s} \tag{B5}$$

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 B3 and B5

$$IDEB_s = N_s \cdot P_s \tag{B6}$$

$$0 \leq N_s \leq 10\,;\; 0 \leq P_s \leq 1\,;\; 0 \leq IDEB_s \leq 10 \tag{B7}$$

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 B3, 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 5$^{\text{th}}$ grade, but not in the last years of middle and high 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 B1.

In line with the baseline results on standardized test scores and passing rates, the only significant effect is found in 6$^{\text{th}}$ 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.[51] As one can see in Figure A10, 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.[52]

Table B1: 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

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

[51] 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.

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

B.2 Net Present Value of Increased Learning

Increased learning is associated with long-term labor market returns, assuming that the accumulation in 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

$$NPV = \sum_{k=20-\alpha}^{N} \frac{\Delta Y \cdot \beta \cdot w}{(1+i)^k} \tag{B8}$$

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.[53] As students enter the labor marker 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 worklife of a student.

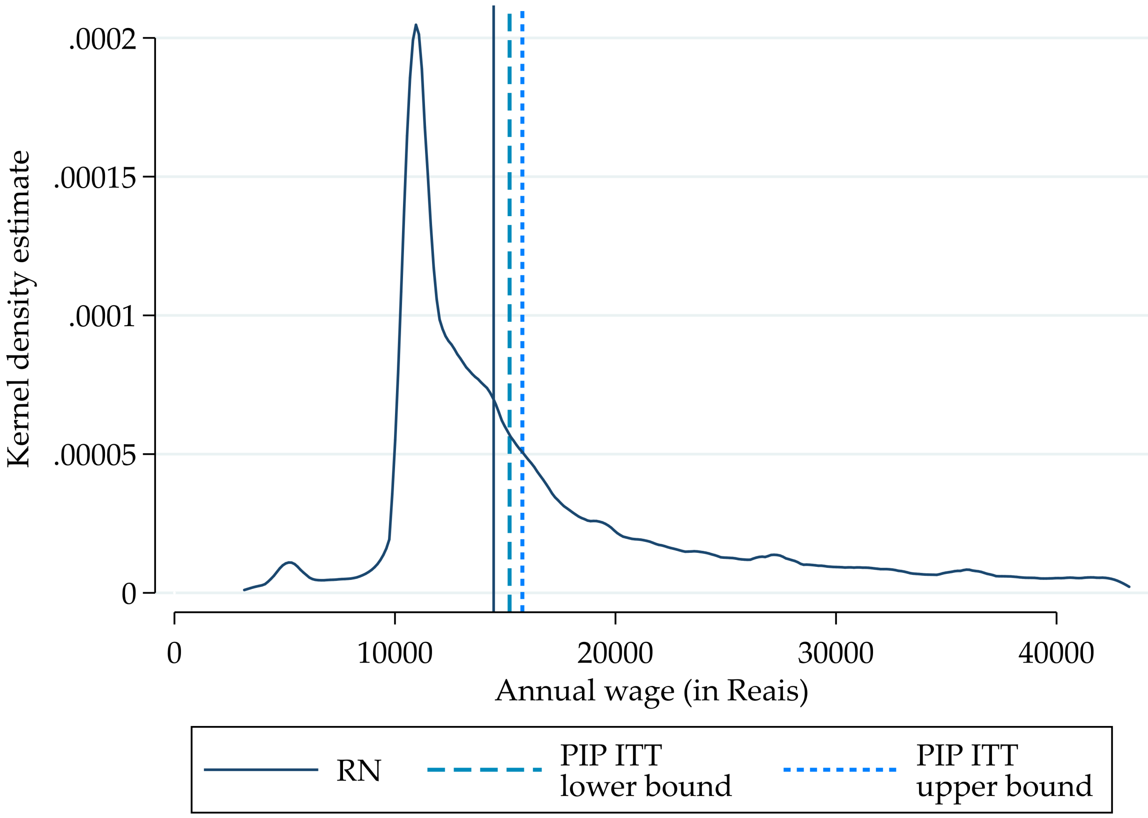
We use the 2016 Annual Social Information Report (*Relação Anual de Informações Sociais* – RAIS) from the Ministry of Labor and Employment to retrieve the average wage in RN (this refers to the formal sector) and, therefore, to estimate the return to education in RN through a conventional Mincerian equation (Mincer, 1974). Namely, the average wage in 2016 was 24,486.48 BRL, i.e., around 6,000 US$. In line with recent estimates by Psacharopoulos and Patrinos (2018), we find the return to one extra year of education in RN to be around 10%. The age of 6$^{\text{th}}$ 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%.

Using our ITT estimates, as computed in Section 4.2, we find that PIP would increase annual wages by 5%. This would mean a shift of the median worker to the 6$^{\text{th}}$, or 7$^{\text{th}}$, decile, respectively, in the wage distribution of RN (Figure B1). Considering the whole worklife, the intervention has

[53] 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.

a predicted NPV between 29,148.97 and 52,468.15 BRL, i.e., 7,287.24 to 13,117.03 US$. This is equivalent to about one average annual Brazil income per capita.

Figure B1: Learning Gains in 6$^{\text{th}}$ Grade Rescaled to Annual Wage



Notes: 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 *Relação Anual de Informações Sociais* (RAIS). N = 801,956.