Journal of Public Economics 234 (2024) 105123

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Teacher-led innovations to improve education outcomes: Experimental evidence from Brazil ☆

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

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

ARTICLE INFO

JEL classification:
H52
I21
M52
O15

Keywords:
Education policy
Lower-secondary education
Autonomy
Mentoring
School resources
Teacher motivation
Achievement
Socio-emotional skills

ABSTRACT

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

1. Introduction

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

Attempts to improve student outcomes often focus on increasing teacher effectiveness due to their central role in the education production function (Chetty et al., 2014; Araujo et al., 2016; Jackson, 2018; Bau and Das, 2020). This goal can be pursued by improving teacher skills and pedagogy and/or by providing (monetary or non-monetary) incentives to strengthen teacher motivation (World Bank, 2018). We present experimental evidence of an education policy in Brazil that provided teachers with support to autonomously design and

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

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

https://doi.org/10.1016/j.jpubeco.2024.105123
Received 31 August 2022; Received in revised form 4 April 2024; Accepted 8 April 2024

implement a local project to tackle their specific issues instead of a centrally defined curriculum ("prescribing solutions"). The program encouraged teachers to propose pedagogical innovations relevant to their specific context, with the goal of improving both student progression and learning outcomes through increased teacher and student engagement.

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

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

The program was implemented by the State Secretariat of Education (SEE) of Rio Grande do Norte (RN) in Brazil. RN consistently scores at the bottom of the Brazilian Education Development Index (*Índice de Desenvolvimento da Educação Básica*—IDEB).[3] While its cornerstone was to provide teachers with autonomy, the program was designed as a comprehensive package, including technical assistance and a limited grant, to support teachers in the design and implementation of the proposed activities. Through seminars and support from a dedicated mentor, teachers developed a diagnostic of their main pedagogical challenges and context-specific innovations to address them. Schools were encouraged to introduce innovative pedagogical projects and expose students to non-curricular competencies and learning opportunities outside the classroom, such as the development of school radio and video production, the setup of theater plays or book fairs, and robotics classes. Mentors complemented local capacity while ensuring close ties with central government, possibly reducing moral hazard concerns associated with the strategic behavior of local staff. Approved proposals were awarded financial support to implement the projects, ranging from about 7500 to 11,000 US dollars (USD),[4] or median USD 139 per student, i.e., 3.6 percent of average annual expenditure per student in Brazil (OECD, 2016).

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

We first show that the program positively impacted student learning and school progression, the outcomes targeted by the program. To assess the program's effects on student learning, we use the state's standardized exam that was introduced in 2016 and extended to grades selected for the impact evaluation. The intent-to-treat (ITT) estimates show no overall impacts, but splitting the analysis by grade reveals substantial impacts in grade 6: a 0.18 SD improvement to math scores and a 0.16 SD improvement to Portuguese scores. Slightly lower impacts are observed on humanities (0.10 SD) and natural sciences (0.12 SD). We estimate the average impact on learning to be equivalent to half an extra year of schooling, or 0.36 years per USD 100 spent.

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

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

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

Third, we leverage the fact that most 6th-grade teachers also teach other grades. Mechanically, we find a similar increase in 7th-grade teacher retention, yet do not find any (positive or negative) spillover impacts on the progression rates of 7th-grade students. This provides suggestive evidence that increased teacher engagement may not necessarily spillover to other grades in the absence of the in-school developed projects.

Taken together, our findings indicate that a teacher-led approach can work, particularly if complemented with other inputs; the targeted

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

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

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

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

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

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

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

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

Finally, while previous experimental evaluations of school grants have shown mixed results, such programs hide a wide degree of heterogeneity in terms of grant size, design features, and decision-making responsibilities within the school.[7] The key novelty of the intervention examined in this paper is that the allocation of funds was bound to pedagogical activities designed and implemented by teachers as opposed to school management or specific items mandated by the central government. The one-time grant was conditional, i.e., the money could not be spent on other school expenses, such as teacher salaries or infrastructure improvements.

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

2. Context and intervention

2.1. Education in Brazil and Rio Grande do Norte

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

These national figures hide a high degree of regional variation. In this paper, we study an education program implemented by the RN state government, one of Brazil's poorest states. In the 2015 national standardized exam,[8] RN state schools scored at the bottom of the learning distribution in both primary and lower-secondary education.[9] The difference in 5th-grade proficiency levels between the average student in RN and the best-performing state is the equivalent of 2.5 years of education.[10] The low level of learning is reflected in the state's progression indicators. In 2015, the average school dropout rate in upper-secondary education was 12.4 percent compared to the national average of 8.8 percent (INEP, 2015). The combination of high dropout rates and low learning outcomes puts RN state schools near the bottom of the Brazilian Index of Development of Basic Education (online Appendix Figure B1).

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

A major constraint to school quality and student achievement in Brazil is principal and teacher turnover, which is around 21 percent nationally (Akhtari et al., 2022). In the RN public school system, 30 percent of teachers leave their schools each year, with little variation across grades (online Appendix Figure B3), potentially disrupting

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

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

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

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

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

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

school operations and compromising personnel collaboration.[11] Using school-level data from INEP, we find that teacher permanence is positively correlated with student passing rates and negatively correlated with age-grade distortion, repetition, and dropout, for both primary and secondary schools (online Appendix Table D1).[12]

2.2. The Pedagogical Innovation Project (PIP)

The Pedagogical Innovation Project (*Projeto de Inovação Pedagógica*—PIP), developed by the RN SEE, aimed at improving both student progression and learning outcomes by increasing child and teacher engagement. The intervention has four main components: (i) a high degree of autonomy for teachers to design and implement a project based on their diagnostic of the context-specific challenges; (ii) the introduction of pedagogical innovations in the classroom; (iii) continuous technical support to teachers during the design and implementation of the project, with the SEE having only an advisory role to assure minimum quality standards; and (iv) a grant specifically earmarked to implement the project.

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

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

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

2.2.1. Project development

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

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

2.2.2. Implementation support and monitoring

Schools with approved proposals are awarded a fixed amount of funding to execute their projects. Schools can only spend the operating funds on inputs directly related to their proposed project. The grant amount depends on the number of classes included in the project and ranges from BRL 30,000 to 45,000, i.e., USD 7576 to 11,364 (online Appendix Figure B4). The median transfer per enrolled student was BRL 555.55, the equivalent of USD 139, which represents about 3.6 percent of average annual expenditure per student in Brazil (OECD, 2016).

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

2.2.3. Characteristics of sub-projects

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

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

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

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

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

3. Experimental design and data

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

3.1. Experimental design

To ensure enough operational capacity, only a sub-sample of schools was selected to participate each year. To determine the pool of eligible schools for implementation during the 2016 school year, three filters were applied. First, only schools that would not change principals between the 2015 and 2016 school year were included to ensure continued buy-in for the prepared projects. State legislation requires directors to change schools every two years, resulting in about half the schools changing directors each year.[14] Second, the 2016 edition

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

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

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

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

targeted the final grade of primary education (5th grade), the first grade of lower-secondary education (6th grade), and the first grade of upper-secondary education (10th grade).[15] Only schools offering at least one of those three grades were considered. Finally, schools that participate in the Federal program ProEMI (*Ensino Médio Inovador)* were excluded.[16] As a result, of the 639 state schools, 299 were eligible to receive the PIP program in 2016.

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

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

The randomization resulted in 130 eligible schools in the treatment group and 169 in the control group (Panel A in online Appendix Table B1). All 130 selected schools were invited to the workshops held in the final months of the 2015 school year. The randomization was performed using the 2015 school census. After the start of the 2016 school year, a few schools had closed or no longer offered the grade that had been selected for the intervention.[18] This leaves us with a final sample of 280 schools effectively allocated to the experiment at the beginning of the 2016 school year (Panel B in online Appendix Table B1)—126 in the treatment group and 154 in the control group. The geographical distribution and treatment assignment of these schools are shown in the online Appendix Figure B5. Across the selected grades in each school, 19,899 students were included in the experiment—9432 in treated schools and 10,467 in control schools (Panel C in online Appendix Table B1).

3.2. Data

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

Administrative Data. We use administrative data from both the state's education monitoring system and the annual national school census to obtain school, teacher, and student characteristics and progression. The state's monitoring system, the *Sistema Integrado de Gestão da Educação* (SIGEduc) portal, provides data on passing, dropout, and repetition rates at the grade level.[19] The school census is carried out on an annual basis by the *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP) of the Brazilian Ministry of Education. It contains information on overall school characteristics, such as location, presence of a library, science lab, and internet, as well as information on the number of teachers, students, and classes.[20] The census also allows us to track individual teachers and students over time, even if they move to other schools within the state.[21] Where possible, the analysis of the results uses both sources.

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

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

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

In addition to the standardized measure of learning, we obtained students' final-year GPAs from *SIGEduc*. These are cross-subject scores, as evaluated by their teachers, on a 10-point scale. Importantly for the attrition analysis featured in the next sub-section, this variable is reported for each single student in our experimental sample and is not restricted to the sample who remain in schools until the end of the year and so participates in the standardized assessment.

Socio-Emotional Skills. To analyze the impact on socio-emotional learning, we measure the Big Five personality traits (neuroticism, extroversion, conscientiousness, agreeableness, and openness). We use an off-the-shelf self-reported test developed and adapted to younger students in Brazil by the *Instituto Ayrton Senna*.[23] This test, and its equivalent, are widely used in the literature to assess socio-emotional skills.[24,25] The test was administered at the end of the 2016 school year

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

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

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

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

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

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

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

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

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

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

[25] Research has shown that individuals with the same level of a trait may assess themselves at very different levels on a Likert scale (Primi et al.,

to the grade that entered the randomization (see Section 3.1). In case a school had multiple classes in the same grade, one class was randomly chosen for the test.

3.3. Validity of the experiment

Balance. To examine whether the randomization resulted in balanced samples across control and treatment groups, we compare observable characteristics prior to the roll-out of the program. Online Appendix Table C1 shows several characteristics at the school, grade, teacher, and student levels, including some of the key outcomes of the intervention, such as repetition and dropout rates. For grade, teacher, and student comparisons, we only consider the classes in the eligible grade for that school (see description in Section 3.1). Columns (2) and (4) show the means in the treatment and control groups. In column (5), we report both standard *p*-values based on *t*-test of differences in the means and *p*-values computed using randomization inference. Generally, we find no statistical differences when comparing the treatment and control groups. A joint significance test of school and student characteristics confirms that these variables do not jointly predict treatment assignment (F-stat of 0.69 and 1.76, respectively).

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

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

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

4. Empirical strategy and results

4.1. Empirical strategy

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

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

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

There are three potential sources of non-compliance with the treatment. First, not all assigned schools received all components of the program, as discussed in Section 3.1. Second, not all teachers at an assigned school participated in the program. Finally, while participating teachers were supposed to apply the pedagogical projects in all their classes, there might be variation in treatment exposure/intensity due to variation in the number and types of teachers participating. To account for this, we include all schools, teachers, and students of the assigned grade, per their original assignment. Therefore, the parameter β identifies the ITT effect. To estimate treatment-on-the-treated (TOT) effects, we use the school random assignment as an instrumental variable to account for non-random variation in teacher participation in the program.

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

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

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

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

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

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

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

[31] The blocked difference-in-means approach uses strata sizes, instead of fixed effects, to weight the treatment effects estimates within each stratum. Gibbons et al. (2018) show that, in the presence of heterogeneous

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

4.2. Results

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

4.2.1. Learning outcomes

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

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

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

4.2.2. Student progression

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

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

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

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

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

The reduction in 6th-grade repetition might have long-term implications for students' years of education and likelihood of completing school. To explore how much improving progression may affect students' school careers, we track all RN students who were in 6th grade in 2011 up to 2017 using school census data. We find that students who were promoted in 6th grade in 2011 are 40 pp more likely to be in school in 2017 than students who were retained in 2011 (online Appendix Figure B6a). Similarly, after six years, they have completed 2.34 more years of schooling (online Appendix Figure B6b). We quantify the correlation between retention in 6th grade and schooling outcomes by estimating an OLS regression of dropout and completed years of schooling on grade repetition.[34] We find that failing 6th grade is associated

treatment effects, fixed effects estimates are generally not a consistent estimator of the average treatment effect. Therefore, they propose IWE and RWE as alternatives to recover such parameter.

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

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

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

Table 1
Impact on student learning.

	(1)	(2)	(3)	(4)	(5)
	Average	Math	Portuguese	Human sciences	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	12 760	11 366	11 365	10 885	10 879
Number of clusters	264	264	264	264	264
Mean dep. var. control group	184.052	172.693	190.234	186.477	185.329
SD dep. var. control group	41.081	46.528	52.637	49.517	42.864
5th grade—Primary schools					
Treatment	−0.068	−0.067	−0.091	−0.070	−0.074
	(0.087)	(0.097)	(0.091)	(0.087)	(0.084)
	[0.584]	[0.641]	[0.407]	[0.531]	[0.417]
Number of observations	3179	2885	2885	2977	2978
Number of clusters	92	92	92	92	92
Mean dep. var. control group	157.452	157.540	173.368	154.288	149.499
SD dep. var. control group	36.022	43.798	60.456	37.359	28.700
6th grade—Lower secondary schools					
Treatment	0.146**	0.177**	0.158**	0.103*	0.123**
	(0.061)	(0.073)	(0.075)	(0.054)	(0.062)
	[0.034]	[0.026]	[0.057]	[0.070]	[0.054]
Number of observations	4511	4014	4013	4134	4131
Number of clusters	99	99	99	99	99
Mean dep. var. control group	162.845	151.930	172.451	160.075	170.685
SD dep. var. control group	31.523	42.024	47.502	35.775	35.164
10th grade—Upper secondary schools					
Treatment	−0.011	−0.015	−0.016	−0.026	0.051
	(0.078)	(0.088)	(0.112)	(0.062)	(0.053)
	[0.884]	[0.847]	[0.878]	[0.610]	[0.417]
Number of observations	5070	4467	4467	3774	3770
Number of clusters	73	73	73	73	73
Mean dep. var. control group	215.446	198.009	214.086	233.701	223.680
SD dep. var. control group	26.923	38.838	41.371	26.369	23.650

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

with a 21 pp higher likelihood of school dropout after six years, and a reduction of 1.7 years of completed schooling (online Appendix Table D5). Taken at face value, our estimates provide suggestive evidence that the reduction of 23 percent (or 7 pp) in repetition rate caused by the PIP might contribute to substantially reducing school dropout (by 4.83 pp) and increasing years of schooling (by 0.4 extra years) of the treated cohort of 6th graders.

4.2.3. Robustness to attrition and changes in sample composition

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

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

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

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

5. Potential mechanisms and heterogeneous results

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

[35] This pattern is observed in each grade.

Table 2
Impact on student progression rates.

	Grade level				*Student level*			
	(1) All	(2) 5th	(3) 6th	(4) 10th	(5) All	(6) 5th	(7) 6th	(8) 10th
	Passing							
Treatment	4.70**	2.44	8.46**	2.46	4.51**	1.29	7.00**	4.25
	(1.83)	(2.55)	(3.30)	(3.61)	(2.21)	(2.65)	(3.10)	(4.13)
		[0.548]	[0.032]	[0.548]		[0.520]	[0.016]	[0.335]
Number of observations	277	95	104	78	17 276	3629	5490	8157
Number of clusters					277	95	104	78
Mean dep. var. control group	70.97	83.55	63.56	66.22	59.91	79.60	58.73	52.81
SD dep. var. control group	18.04	13.64	17.05	16.23	49.01	40.31	49.24	49.93
	Dropout							
Treatment	−0.20	−0.16	−1.61	1.60	−0.85	0.26	−4.35**	1.13
	(0.83)	(0.79)	(1.27)	(2.21)	(1.39)	(1.38)	(1.82)	(2.67)
		[0.845]	[0.492]	[0.714]		[0.822]	[0.009]	[0.822]
Number of observations	277	95	104	78	17 290	3637	5494	8159
Number of clusters					277	95	104	78
Mean dep. var. control group	6.19	2.09	6.84	10.17	16.83	8.19	13.55	22.40
SD dep. var. control group	7.96	3.87	7.15	10.17	37.42	27.43	34.23	41.70
	Repetition							
Treatment	−4.49***	−2.28	−6.85**	−4.06	−3.66**	−1.55	−2.65	−5.38*
	(1.70)	(2.38)	(2.91)	(3.61)	(1.69)	(1.87)	(2.81)	(2.97)
		[0.445]	[0.053]	[0.445]		[0.390]	[0.390]	[0.066]
Number of observations	277	95	104	78	17 276	3629	5490	8157
Number of clusters					277	95	104	78
Mean dep. var. control group	22.84	14.37	29.59	23.61	23.25	12.20	27.72	24.78
SD dep. var. control group	15.27	12.86	14.91	13.71	42.25	32.74	44.77	43.18

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

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

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

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

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

5.1. Impact on teacher engagement

We proxy teachers' engagement by their decision to remain in the school, measured as teacher retention (see Section 3.2).[36] Low teacher retention might affect students' achievement and motivation, particularly in 6th grade. Mentors stated that 6th graders face significant shocks when transitioning between levels of education. In primary school, students have a single teacher, which allows for a close student–teacher relationship. These ties are weaker in lower secondary education, as students have multiple teachers (at least five in our sample). The potential negative impact of this transition is well documented in the United States (Bedard and Do, 2005; Cook et al., 2008; Hanewald, 2013) and has been recently investigated in Brazil (Santos et al., 2017). The latter study evaluates the impact of a pilot in municipal schools in Rio de Janeiro, Brazil, which expanded primary schools to include grade 6. They find that having 6th grade in the primary school increases learning by 0.16 SD and they provide suggestive evidence that strengthened student–teacher relationships mediated some of the effects on learning. This might imply that enhancing teacher retention in 6th grade could be particularly effective in improving student outcomes, in line with recent evidence from Ronfeldt et al. (2013), Fagernäs and Pelkonen (2020), Zeitlin (2021), and Akhtari et al. (2022)—the latter in the Brazilian context.

Pooling all grades, the ITT estimate on teacher retention is positive but not statistically significant (column 1 in Panel 1 of Table 3). Splitting the analysis by grade, we find that teacher retention increased by 6.4 pp in grade 6, i.e., a 20.7 percent increase in teacher retention

[36] We note that there are other channels, besides enhanced autonomy and motivation, that we do not directly observe through which the program might affect teacher retention. For example, if the program was perceived as a signal that the government has increased the degree of resources that will be provided to the school. We explore the role of the grant in the mechanism section.

over the control mean of 30.9 percent (column 3).[37,38] We also recover TOT estimates for participating teachers as the program did not require all teachers to participate and the decision for which and how many teachers to include was at the discretion of the schools. Since primary schools have class-specific teachers, while secondary schools have subject-specific teachers, teacher participation was mechanically higher in 5th grade. We find similar rates of participation in 6th and 10th grades, where roughly one third of the teachers complied with treatment assignment (Panel 1 of online Appendix Table D6).[39] In 6th grade, the TOT estimate for teacher retention is equal to 19 pp, or 61 percent over the control mean (Panel 2 of online Appendix Table D6). Moreover, we find that observable differences between participant and non-participant teachers are small in magnitude and mostly not distinguishable from zero (online Appendix Table B2). This suggests that the differential selection across grades along these characteristics is not likely driving the heterogeneous results.

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

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

5.2. Implementation of innovative pedagogical projects

Project overall quality. We first explore whether the heterogeneous results by grade are driven by better-designed projects in grade 6. However, we find no differences across grades in the average quality of proposals, measured by their score (online Appendix Figure B7). Moreover, these scores do not seem to have any predictive power on either the rate of project implementation or student outcomes, such as progression and learning (online Appendix Table D7 and D8, respectively).

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

5.3. Role of teacher retention and project implementation

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

Further, we explore whether the observed patterns of impacts on teacher retention, being concentrated in schools with low teacher retention at baseline, and implementation, are also observed for child outcomes. In grade 6, we observe no difference in in-school implementation by baseline levels of teacher retention (online Appendix Figure B8), while for grade 10 implementation is substantially worse in schools with low teacher retention at baseline. Impacts on progression do not differ by baseline levels of teacher retention, while learning gains are concentrated among schools with low retention at baseline, where teachers are most affected (first panel of online Appendix Table D13). This is in line with strong, positive correlations between student learning outcomes and teacher retention in the control group, while correlations with student progression are more mixed. This might suggest that teacher retention, and engagement, may be critical to generating learning gains through the projects. In grade 10 schools with low teacher retention at baseline, teachers were impacted but implementation of projects was particularly poor, possibly explaining the lack of impacts on learning (second panel of online Appendix Table D13).

5.4. Student–teacher interactions

Changes in teacher–student interactions. We explore descriptively whether learning results are driven by the educational content of the projects or by changes in pedagogy and student–teacher interactions, which can also be induced by increased teacher engagement. While 70 percent of projects implemented in 6th grade were focused on reading

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

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

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

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

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

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

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

Table 3
Impact on teacher retention.

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

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

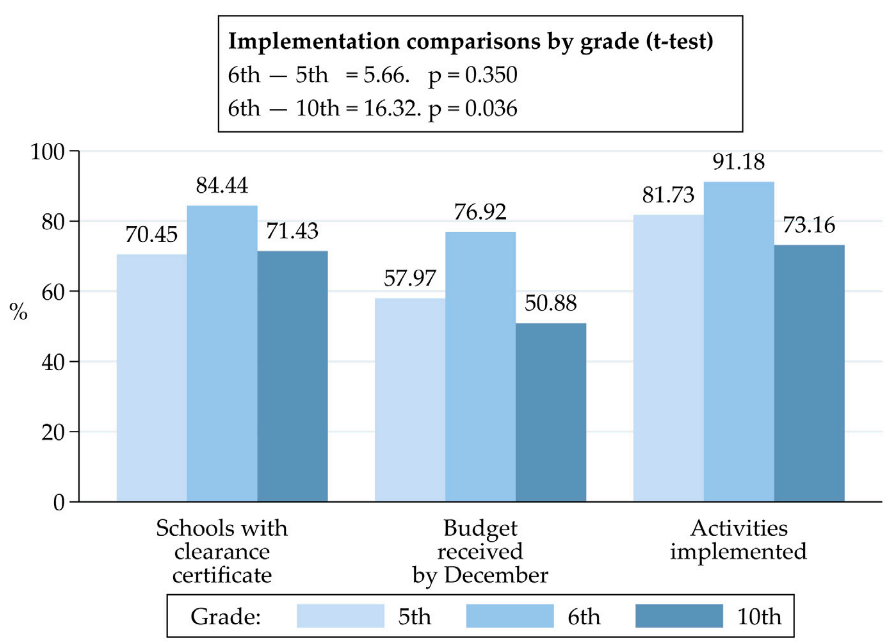

Fig. 1. Implementation by grade. *Note:* The bars show the unconditional mean of the three implementation measures defined in Section 5.5. Comparisons by grade are done on the third measure using region fixed effects and robust standard errors. P-values based on standard t-test. 'Planned activities implemented' are defined as the ratio of the number of activities that were implemented over the number of planned activities described in the work plan. Data are from the State Secretariat of Education (SEE). Sample: schools in the treatment group.

and literacy (online Appendix Figure A1b), we find ITT effects of similar magnitude in all subjects tested. Moreover, the standardized assessment was not high-stakes for students, hence we do not expect teachers to have content tailored to the exam. At the same time, the correlation between scores on different subjects is not particularly high, ranging from 0.34 (human and natural sciences) to 0.57 (math and Portuguese). This limits the likelihood that the impact is explained by the curricular content of the activities only.

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

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

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

5.5. Role of complementary program components

Complementing management capacity. Financial resources do not seem to be the main binding constraint for the procurement of school supplies and implementation of new pedagogical activities in our context. Data on disbursement rates of public funds, such as *Programa Dinheiro Direto na Escola* (PDDE) run by *Fundo Nacional de Desenvolvimento da Educação* (FNDE) reveal that, on average, only 30.5 percent of federal funds assigned to schools to invest in infrastructure and pedagogy are spent.[46] Rather than solving budgetary scarcity, the program might have prompted schools to overcome administrative bottlenecks in access to school funding, for instance by obtaining the clearance certificate. This might have led to increased spending across the board, which in turn supported principals to spend available public funds more effectively. Online Appendix Table D15 shows that this was the case for schools participating in PIP with their 6th grade: treated schools disbursed a 60 pp higher share of PDDE funds compared to the control in the year following the intervention. We do not observe this in the school year we evaluate, but this does suggest that the provision of technical support in project management and procurement is important and can make a substantial difference in the implementation of planned activities.

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

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

5.6. Heterogeneity in other grade characteristics

We explore whether heterogeneity in baseline grade characteristics might explain the differential impact in 6th grade. We start by noticing that the state administers the universe of schools offering upper-secondary education and only a smaller share of schools offering lower-secondary education. We compare observed characteristics across grades to assess whether the 6th grade state schools are different from schools offering either the 5th or 10th grades in the study sample.

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

6. Policy analysis

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

Quality of Education. We use the ITT estimates to compute the counterfactual distribution of two national quality of education indicators. First, online Appendix Figure F1 shows that if students retain their learning gains over time, as measured by SAEB scores, the impact of

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

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

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

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

Table 4
Impact on Socio-Emotional Skills.

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

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

the PIP would close half of the knowledge gap between RN and the country's average by the end of grade 9. Second, combining impacts on progression and learning suggests that the PIP would help RN state schools move upwards in the IDEB ranking by at least two positions (online Appendix Figure F2).[48] This strategy is described in more detail in the online Appendix F.1.

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

Using the ITT effects of the PIP on learning as being approximately equal to 0.5 extra years of schooling, a back-of-the-envelope calculation suggests an NPV on future earnings of BRL 29,148.97 or USD 7287.24. The second channel is through the increase in student years of schooling through a reduction in repetition which we estimate leads to about 0.4 extra years of schooling, with an NPV on future earnings of BRL 23,319.18 (or USD 5829.79). The full effect on expected earnings would then range from USD 7000 to 13,000, or 28 to 52 times the annual Brazilian minimum wage. The data and methodology used for the calculations are described in online Appendix F.2.

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

7. Conclusion

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

We found a positive impact of 6.6 percent on progression rates, with the effects being concentrated in grade 6 (12 to 13 percent increase over the control mean). We detected a meaningful positive impact on learning outcomes for 6th graders only. The positive effects on math

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

and reading scores point to an impact equivalent to almost half a year of additional schooling.

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

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

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

Declaration of competing interest

The World Bank i2i fund and the World Bank Brazil Country Office funded the data collection for this study.

I hereby declare that I have no relevant or material financial interests that relate to the research described in this paper.

I am employed at the World Bank, who was responsible for lending funds to the State Secretariat of Rio Grande do Norte, Brazil, to implement the intervention studied in this manuscript.

Caio Piza is a member of the Development Impact Department of the World Bank, which maintains separation from the operations functions of the institution.

Astrid Zwager is a member of the Development Impact Department of the World Bank, which maintains separation from the operations functions of the institution.

Matteo Ruzzante is a PhD candidate in Economics at Northwestern University.

Rafael Dantas is a member of the Development Impact Department of the World Bank, which maintains separation from the operations functions of the institution.

Andre Loureiro is a member of the Education Global Practice of the World Bank and co-led the World Bank operation with the state of Rio Grande do Norte.

The Brazil Country Office Director, Paloma Anos Casero, was given a draft of the manuscript and invited to provide their comments on the draft. The policy of the World Banks research group is that country teams are invited to inform research staff of issues that may be result from the publication of a manuscript, but that the decision of whether and what to publish lies entirely with the researchers. Country teams do not have the right to delay or deny publication.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jpubeco.2024.105123.

References

Akhtari, M., Moreira, D., Trucco, L., 2022. Political turnover, bureaucratic turnover, and the quality of public services. Amer. Econ. Rev. 112 (2), 442–493.

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

Alves, M.T.G., Soares, J.F., Xavier, F.P., 2016. Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais. Revista Brasileira de Sociol. 4 (7), 49–82.

Andrabi, T., Bau, N., Das, J., Karachiwalla, N., Khwaja, A., 2021. Crowding in private quality: The equilibrium effects of public spending in education. Working Paper.

Angrist, N., Djankov, S., Goldberg, P.K., Patrinos, H.A., 2021. Measuring human capital using global learning data. Nature 592 (7854), 403–408.

Araujo, M.C., Carneiro, P., Cruz-Aguayo, Y., Schady, N., 2016. Teacher quality and learning outcomes in kindergarten. Q. J. Econ. 131 (3), 1415–1453.

Bandiera, O., Best, M.C., Khan, A.Q., Prat, A., 2021. The allocation of authority in organizations: A field experiment with bureaucrats. Q. J. Econ. 136 (4), 2195–2242.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From proof of concept to scalable policies: Challenges and solutions, with an application. J. Econ. Perspect. 31 (4), 73–102.

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

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., Singh, N., 2021. Improving police performance in rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. Am. Econ. J.: Econ. Policy 13 (1), 36–66.

Banerjee, A.V., Cole, S., Duflo, E., Linden, L., 2007. Remedying education: Evidence from two randomized experiments in India. Q. J. Econ. 122 (3), 1235–1264.

Bau, N., Das, J., 2020. Teacher value added in a low-income country. Am. Econ. J.: Econ. Policy 12 (1), 62–96.

Beasley, E., Huillery, E., 2017. Willing but unable? Short-term experimental evidence on parent empowerment and school quality. World Bank Econ. Rev. 31 (2), 531–552.

Bedard, K., Do, C., 2005. Are middle schools more effective? The impact of school structure on student outcomes. J. Hum. Resour. 40 (3), 660–682.

Beg, S., Halim, W., Lucas, A.M., Saif, U., 2022. Engaging teachers with technology increased achievement, bypassing teachers did not. Am. Econ. J.: Econ. Policy 14 (2), 61–90.

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

Burgess, R., Hansen, M., Olken, B.A., Potapov, P., Sieber, S., 2012. The political economy of deforestation in the tropics. Q. J. Econ. 127 (4), 1707–1754.

Carneiro, P., Koussihouèdé, O., Lahire, N., Meghir, C., Mommaerts, C., 2020. School grants and education quality: Experimental evidence from Senegal. Economica 87 (345), 28–51.

Carvalho, J.S.d., 2016. O Projeto de Inovação Pedagógica (PIP) e as Práticas Inovadoras dos Professores da Rede Estadual do Ensino Médio do RN. (Master's thesis). Universidade Estadual do Rio Grande do Norte.

Cassar, L., Meier, S., 2018. Nonmonetary incentives and the implications of work as a source of meaning. J. Econ. Perspect. 32 (3), 215–238.

Chetty, R., Friedman, J.N., Rockoff, J.E., 2014. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. Amer. Econ. Rev. 104 (9), 2633–2679.

Cook, P.J., MacCoun, R., Muschkin, C., Vigdor, J., 2008. The negative impacts of starting middle school in sixth grade. J. Policy Anal. Manage. 27 (1), 104–121.

Cunha, F., Heckman, J., 2007. The technology of skill formation. Amer. Econ. Rev. 97 (2), 31–47.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., Sundararaman, V., 2013. School inputs, household substitution, and test scores. Am. Econ. J.: Appl. Econ. 5 (2), 29–57.

Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. J. Personal. Soc. Psychol. 18 (1), 105.

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

Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2018. The value of regulatory discretion: Estimates from environmental inspections in India. Econometrica 86 (6), 2123–2160.

Evans, D.K., Yuan, F., 2019. Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms. Policy Research Working Paper No. 8752, The World Bank, Washington, DC.
Fagernäs, S., Pelkonen, P., 2020. Teachers, electoral cycles, and learning in India. J. Hum. Resour. 55 (2), 699–732.
Firpo, S., Foguel, M.N., Jales, H., 2020. Balancing tests in stratified randomized controlled trials: A cautionary note. Econom. Lett. 186, 108771.
Firpo, S., Fortin, N.M., Lemieux, T., 2009. Unconditional quantile regressions. Econometrica 77 (3), 953–973.
Gibbons, C.E., Serrato, J.C.S., Urbancic, M.B., 2018. Broken or fixed effects? J. Econom. Methods 8 (1).
Gilligan, D.O., Karachiwalla, N., Kasirye, I., Lucas, A.M., Neal, D., 2022. Educator incentives and educational triage in rural primary schools. J. Hum. Resour. 57 (1), 79–111.
Glewwe, P., Kremer, M., Moulin, S., 2009. Many children left behind? Textbooks and test scores in Kenya. Am. Econ. J.: Appl. Econ. 1 (1), 112–135.
Glewwe, P., Muralidharan, K., 2016. Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In: Handbook of the Economics of Education. vol. 5, Elsevier, pp. 653–743.
Gray-Lobe, G., Keats, A., Kremer, M., Mbiti, I., Ozier, O.W., 2022. Can Education Be Standardized? Evidence from Kenya. Becker Friedman Institute for Economics Working Paper, Evidence from Kenya (June 5, 2022). University of Chicago.
Hanewald, R., 2013. Transition between primary and secondary school: Why it is important and how it can be supported.. Aust. J. Teach. Educ. 38 (1), n1.
Hanushek, E.A., Woessmann, L., 2008. The role of cognitive skills in economic development. J. Econ. Lit. 46 (3), 607–668.
Instituto Brasileiro de Geografia e Estatística (IGBE), 2018. Pesquisa nacional por amostra de domicílios (PNAD) contínua. Rio de Janeiro.
Imbens, G.W., Manski, C.F., 2004. Confidence intervals for partially identified parameters. Econometrica 72 (6), 1845–1857.
Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 2015. Censo escolar. Ministério da Educação: Brasília.
Ivcevic, Z., Brackett, M., 2014. Predicting school success: Comparing conscientiousness, grit, and emotion regulation ability. J. Res. Personal. 52, 29–36.
Jackson, C.K., 2018. What do test scores miss? The importance of teacher effects on non–test score outcomes. J. Polit. Econ. 126 (5), 2072–2107.
Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B., Borghans, L., 2014. Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. NBER Working Paper No. 20749, National Bureau of Economic Research.
Laajaj, R., Macours, K., 2019. Measuring skills in developing countries. J. Hum. Resour. 1018–9805R1.
Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Rev. Econom. Stud. 76 (3), 1071–1102.
Marinelli, H.A., Berlinski, S., Busso, M., 2021. Remedial education: Evidence from a sequence of experiments in Colombia. J. Hum. Resour..
Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., Rajani, R., 2019. Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. Q. J. Econ. 134 (3), 1627–1673.
McEwan, P.J., 2015. Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. Rev. Educ. Res. 85 (3), 353–394.
Muralidharan, K., Singh, A., Ganimian, A.J., 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. Amer. Econ. Rev. 109 (4), 1426–1460.
Organisation for Economic Co-operation and Development (OECD), 2015. PISA: Programme for International Student Assessment.
Organisation for Economic Co-operation and Development (OECD), 2016. Education at a glance 2016. Paris, OECD Publishing, p. 508.
Oliveira, L.F.B., Terra, R., et al., 2016. Impacto do programa mais educação em indicadores educacionais. IPC-IG Working Paper n. 147, International Policy Centre for Inclusive Growth, Brasília.
Poropat, A.E., 2009. A meta-analysis of the five-factor model of personality and academic performance. Psychol. Bull. 135 (2), 322.
Primi, R., Zanos, C., Santos, D., De Fruyt, F., John, O.P., 2016. Anchoring vignettes can they make adolescent self-reports of social-emotional skills more reliable, discriminant, and criterion-valid? Eur. J. Psychol. Assess. 32 (1).
Rasul, I., Rogger, D., 2018. Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service. Econ. J. 128 (608), 413–446.
Rasul, I., Rogger, D., Williams, M.J., 2021. Management, organizational performance, and task clarity: Evidence from ghana's civil service. J. Public Adm. Res. Theory 31 (2), 259–277.
Rogger, D., Somani, R., 2018. Hierarchy and Information. Policy Research Working Paper No. 8595, The World Bank, Washington, DC.
Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73 (4), 1237–1282.
Romano, J.P., Wolf, M., 2016. Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. Statist. Probab. Lett. 113, 38–40.
Ronfeldt, M., Loeb, S., Wyckoff, J., 2013. How teacher turnover harms student achievement. Am. Educ. Res. J. 50 (1), 4–36.
Santos, D.D., Sorzafave, L.G., Nicolella, A.C., Sant'Anna, E.G., 2017. Mais é menos? O impacto do projeto 6° ano experimental–SME/RJ. Estudos em Avaliação Educ. 28 (69), 718–747.
United Nations Educational, Scientific and Cultural Organization (UNESCO), 2020. UNESCO institute for statistics, national monitoring. Paris.
World Bank, 2018. World Development Report 2018: Learning to Realize Education's Promise. Washington, DC, The World Bank.
Zeitlin, A., 2021. Teacher turnover in rwanda. J. Afr. Econ. 30 (1), 81–102.
Zuckerman, M., Porac, J., Lathin, D., Deci, E.L., 1978. On the importance of self-determination for intrinsically-motivated behavior. Pers. Soc. Psychol. Bull. 4 (3), 443–446.