

CAMBRIDGE WORKING PAPERS IN ECONOMICS

The Impact of Cost-Effective Management Practices on Student Learning: Evidence from a Large-Scale Randomised Field Experiment

Felipe Galvão Tiago
Puccioni Cavalcanti

Abstract

Causal evidence on the effectiveness of management in education is limited and ambiguous. In this study, we investigate how cost-effective management practices boost student learning through a randomised field experiment conducted with 31,760 students from 80 grade 1–9 public schools in the municipality of Rio de Janeiro, Brazil. The experiment intervention, delivered by municipal servants exclusively to school managers, involved one-to-one coaching and on-the-job training focused on implementing the World Management Survey (WMS)’s “23 best management practices” for the education sector. We also conducted two double-blind, in-depth management surveys, one prior to and one following the programme implementation, to evaluate precisely the quality of the management of the schools. The surveys were based on the WMS methodology. After two years, the estimated average treatment effects were 0.928 (0.260) SD for school management, 0.226 (0.059) SD for reading, and 0.237 (0.059) SD for mathematics. Instrumental variable estimates indicate that a one-point improvement in school management (on a 1–5 scale) led to gains of 0.680 (0.245) SD in reading and 0.714 (0.265) SD in mathematics. Students in schools achieving a one-point management improvement were more than two academic years of learning ahead of peers in untreated schools. We present causal estimates amongst the largest in the education intervention literature based on a programme that costs only \$15.22 (PPP-adjusted) per student per year. The programme can be applied to any school and has expanded in Brazil.

Reference Details

2515	Cambridge Working Papers in Economics
Published	27 March 2025
JEL Codes	C93, H83, I20, J24 and M10
Website	https://www.econ.cam.ac.uk/publications/cwpe

THE IMPACT OF COST-EFFECTIVE MANAGEMENT PRACTICES ON STUDENT LEARNING: EVIDENCE FROM A LARGE-SCALE RANDOMISED FIELD EXPERIMENT*

Felipe Galvão Puccioni, Tiago Cavalcanti[†]

Causal evidence on the effectiveness of management in education is limited and ambiguous. In this study, we investigate how cost-effective management practices boost student learning through a randomised field experiment conducted with 31,760 students from 80 grade 1–9 public schools in the municipality of Rio de Janeiro, Brazil. The experiment intervention, delivered by municipal servants exclusively to school managers, involved one-to-one coaching and on-the-job training focused on implementing the World Management Survey (WMS)’s “23 best management practices” for the education sector. We also conducted two double-blind, in-depth management surveys, one prior to and one following the programme implementation, to evaluate precisely the quality of the management of the schools. The surveys were based on the WMS methodology. After two years, the estimated average treatment effects were 0.928 (0.260) SD for school management, 0.226 (0.059) SD for reading, and 0.237 (0.059) SD for mathematics. Instrumental variable estimates indicate that a one-point improvement in school management (on a 1–5 scale) led to gains of 0.680 (0.245) SD in reading and 0.714 (0.265) SD in mathematics. Students in schools achieving a one-point management improvement were more than two academic years of learning ahead of peers in untreated schools. We present causal estimates amongst the largest in the education intervention literature based on a programme that costs only \$15.22 (PPP-adjusted) per student per year. The programme can be applied to any school and has expanded in Brazil.

JEL Codes: C93, H83, I20, J24 and M10.

*We have benefited from comments from Razan Amine, Ben Arold, Ricardo Paes de Barros, Chris Rauh, André Souza and participants at presentations at numerous institutions and conferences. We are grateful to Luiz Antonio Guaraná and Renan Ferreirinha for their assistance and support. This research would not have been possible without the collaboration of the Rio de Janeiro Municipal Court of Accounts (TCMRio), Rio de Janeiro City Hall, and the Rio de Janeiro Municipal Education Secretariat (SMERJ). We extend our gratitude to the programme’s implementation team and all the school managers who participated in this project. We take responsibility for any remaining errors.

[†]Address: University of Cambridge, Faculty of Economics, Sidgwick Avenue, CB3 9DD, United Kingdom. Telephone number: +44 1 223 335262. E-mail: tvdvc2@cam.ac.uk.

I. INTRODUCTION

Educated people are generally healthier, live longer, and earn more than those with less education (Card 1999; Krueger 2003; Chetty, Friedman, and Rockoff 2014; Davies et al. 2018). Moreover, evidence suggests that education influences occupational status, success in the labour market, marital happiness, children’s achievements, decision-making, and many other non-pecuniary aspects of life (Oreopoulos and Salvanes 2011). Therefore, it is unsurprising that education has been recognised as a fundamental human right since the Universal Declaration of Human Rights 1948. At the aggregate level, economic theory suggests that human capital is a key determinant of long-term growth (Lucas 1988; Becker, Murphy, and Tamura 1990), and broad empirical evidence from a number of countries corroborates a positive relationship between education and economic growth (Barro 2001).

Students in most developing countries and underprivileged ones in the developed world are schooled¹ but not necessarily educated (Pritchett 2013; World Bank 2018; Angrist et al. 2021). Nearly all developing countries are many years of learning behind developed countries despite similar years of schooling (Pritchett 2013; World Bank 2018; Angrist et al. 2021). This learning inequality also exists within developed countries. The average difference between underprivileged pupils and more privileged ones with similar years of schooling in developed countries is above 100 PISA score points as shown by the PISA index of economic, social and cultural status (ESCS) (OECD 2023). Therefore, discussing how to enhance underprivileged children’s learning is of utmost relevance.

While several factors may partially influence students’ educational performance—such as family background, cultural aspects, and educational investment—how the education resources are utilised through public policies could play a decisive role (Angrist et al. 2010; Abdulkadiroğlu et al. 2011; Dobbie and Fryer 2011; Angrist et al. 2012; Curto and Fryer 2014; Fryer 2014; Tavares 2015; Fryer 2017a; Barros et al. 2019; Bloom et al. 2020;

¹The Millennium Development Goal of achieving universal primary education set in 2000 by the United Nations was considered almost met by 2015, with 91 % of children from developing countries enrolled in primary schools. See <https://www.mdgmonitor.org/mdg-2-achieve-universal-primary-education/>.

Barros et al. 2021; Beg, Fitzpatrick, and Lucas 2023; Alan and Kubilay 2025). This paper investigates how “the best management practices” impact students’ learning in Brazilian public schools. To this end, we designed and implemented the first randomised field experiment to test the causal effects of “the 23 best management practices for schools”—which have been extensively discussed in the literature and outlined in Bloom et al. 2015—on students’ learning. The 23 best management practices for schools are described in Appendix D.

To be the treatment arm of the experiment, we develop a novel real-world programme—scalable and at an annual per-student cost of only US\$ 15.22, adjusted by purchasing power parity (PPP)—to provide schools with the 23 best managerial practices. We named the intervention the Science and Management for Education Programme (SMEP). SMEP is delivered exclusively to school managers, with no contact between the implementation team and teachers or students. The programme involved one-to-one coaching sessions and on-the-job training focusing on the 23 best management practices. The intervention did not provide financial incentives, increase school or teaching time, or change systems and personnel.

The experimental sample was formed by 31,760 students from 80 schools that were randomly sampled (in September 2021)² from the population of 992 grade 1–9 public schools in Rio de Janeiro, Brazil, home to one of Latin America’s largest public education networks.³ Using a pair-matching procedure, 40 schools were randomly assigned to the treatment group and 40 to the control group.

Unlike many field experiments, we did not contract external consultants to deliver the intervention. Rather, we formed the programme implementation team by partnering with Brazil’s Rio de Janeiro Municipal Court of Accounts (TCMRio) and the Rio de Janeiro Municipal Education Secretariat (SMERJ), which allocated six civil servants to the project.

To evaluate the quality of management with respect to the 23 specific practices tar-

²Muralidharan and Niehaus (2017) found that in the top five economic journals published between 2001 and 2016, only 31% of the randomised controlled trials (RCTs) drew units from a larger population.

³See <https://dl.acm.org/doi/abs/10.1145/3614321.3614379>.

geted by the intervention, we conducted two double-blind management surveys: one before the programme implementation started in September 2021 and another just after in December 2023. The surveys were based on the World Management Survey (WMS) methodology. The surveys were double-blind since school managers did not know that their schools were being evaluated, and the interviewers were unaware of the schools' information. The scores for each practice ranged from 1 (worst management level) to 5 (best management level). All managers and schools were scored twice in both surveys, ensuring that each management practice in each school had two independent scores provided by two different raters. The high agreement between the raters in both surveys suggests that the management scores given to schools were not subjective views of the coders but accurately reflected the management level of schools. To our knowledge, this is the first management experiment to have generated systematic and detailed management data through robust surveys that followed the consolidated WMS survey methodology.

Our first finding is that our programme substantially impacted the average management level of the schools. The estimated average treatment effect (ATE) of the programme on overall school management, measured after two years of treatment, was 0.928 (0.260) standard deviations (SD), which was statistically significant at the 1% level.

Second, we find that the intervention not only significantly affected school management but also had a nontrivial first-order impact on student learning. The ATE estimates of SMEP on students' educational outcomes were 0.226 (0.059) SD for reading and 0.237 (0.059) SD for mathematics, which were statistically significant at the 1% level. Our measures of students' learning were derived from standardised independent assessments that evaluate all students from the first to the ninth grades. After two years of intervention, the students in the treatment group were $\frac{3}{4}$ of a school year of learning (reading and mathematics) ahead of those in the control group. Treatment schools were, therefore, 37.5% more productive in providing reading and mathematics instruction to their students than non-treatment schools. Furthermore, treatment significantly affected student absenteeism by reducing it by 19%.

Our heterogeneous analysis demonstrated that the differences in the programme's

impact across various subgroups—race, gender, poverty level, school segments (I: grades 1-5 schools vs II: grades 6-9 schools), school size (number of students), or students’ baseline performance groups—were minimal and not statistically significant.

Third, we conducted a treatment intensity analysis, which uses the random assignment as an instrumental variable (IV), to demonstrate that 1 score-point improvement in the general school management—on a scale that goes from 1, worst management, to 5, best management—impacted students’ educational achievement by 0.680 (0.245) SD for reading and 0.714 (0.265) SD for mathematics. Likewise, we show that a 1-standard-deviation change in the overall school management had a causal effect of 0.247 (0.089) SD for reading and 0.259 (0.096) SD for mathematics. These estimates were statistically significant at the 1% level. Students in high-implementation schools—i.e., those where the treatment improved their management level by around 1 score point—were more than two academic years of learning ahead of students in contrafactual schools. Therefore, these schools were approximately 110% more productive in providing learning to their students after two years of treatment.

This is the first paper to show the causal relationship between management, precisely measured using the renowned WMS methodology, and student learning outcomes. Using a cross-sectional analysis without a causal estimate, Bloom et al. (2015) found that a 1 SD increase in a managerial index similar to ours was associated with an increase of 0.232 SD in student educational outcomes. Our results, therefore, provide validation for their nonexperimental findings.

Related literature. This paper relates to a growing body of economic research examining the relationship between management and productivity across different sectors (Bloom and Van Reenen 2007, 2010; Bloom et al. 2012a,b; Angrist, Pathak, and Walters 2013; Dobbie and Fryer 2013; Bloom et al. 2014, 2015; Bloom, Sadun, and Van Reenen 2016; Bruhn, Karlan, and Schoar 2018; Bloom et al. 2019, 2020; Gosnell, List, and Metcalfe 2020).

This work also relates to the literature on the impact of charter schools on student performance (Angrist et al. 2010; Abdulkadiroğlu et al. 2011; Dobbie and Fryer 2011;

Angrist et al. 2012; Curto and Fryer 2014; Fryer 2014). However, we focus on the impact of low-cost management practices on student learning without increasing school or teaching time; changing teachers, principals or any other school staff; or providing financial incentives.

Fryer (2014) conducted an RCT in which 20 of the lowest-performing public schools in Houston, Texas, were transformed into charter schools. The changes included substituting nearly all principals and half the teachers before the experiment started and exposing the students to at least 20% more teaching time. Although this is an important experiment, it is not possible to disentangle the effect of management practices from the impact of changes in school staff.

The differences between the mentioned charter school literature and our experiment explain the much higher annual cost per student of charter school interventions compared with that of our intervention. For example, Curto and Fryer (2014) showed that attending a School for Educational Evolution and Development (SEED), which combines a “No Excuses” charter model with a 5-day-a-week boarding programme, increased achievement by 0.211 SD in reading and 0.229 SD in mathematics per year. The annual per-student cost of the SEED was USD 51,904.45. In contrast, our treatment delivered higher impacts at an annual per-student cost of USD (PPP) 15.22.

Causal evidence on the effectiveness of specific management practices to improve student learning is still scarce (Tavares 2015; Fryer 2017a; Barros et al. 2019; Hoyos, Ganimian, and Holland 2019; Muralidharan and Singh 2020; Barros et al. 2021; Romero et al. 2022; Beg, Fitzpatrick, and Lucas 2023). Our large-scale randomised field experiment contributes to this relatively rare literature with strong evidence on the impact of cost-effective management practices on educational outcomes while controlling for teaching time and existing staff and excluding performance-based pay.

Our findings solve the conflicting and unclear causal evidence on the effect of management on student learning. On the one hand, Hoyos, Ganimian, and Holland (2019), Muralidharan and Singh (2020), and Romero et al. (2022) showed that management may not affect student learning. Our experiment helps us understand that these null-effect

studies were driven by interventions that a) did not implement the right package of management practices to impact student learning, as not all management practices affect learning, b) failed to ensure that the schools changed their managerial practices due to the intervention since delivering only information or training do not necessarily change organisational management practices, or c) did not change school management enough as small management changes may not be sufficient to increase productivity or to be captured by econometric analysis. Our programme had to change the management of the schools by 0.916 SD to change students' learning by around 0.23 SD.

On the other hand, Tavares (2015), Fryer (2017a), Barros et al. (2019), Barros et al. (2021), and Beg, Fitzpatrick, and Lucas (2023) reported some unclear “positive” evidence regarding the impact of school management on learning. An example of unclear positive results is the randomised controlled trial (RCT) conducted by Fryer (2017a) in which principals from Houston, Texas, received management training for two years. It is the closest study to ours. Despite some similarities, our experiment differs from this principal training experiment in many ways. First, our programme focuses on school managers (i.e., principals, principal deputies, supervisors, and pedagogical coordinators) and not only principals. Second, our team trained and coached the school managers within school facilities, ensuring that the educational units' management practices were changing. Third, our two management surveys made it possible to assess changes in management practices resulting from the intervention and, consequently, the causal impact of such changes on student learning. Fourth, the Houston experiment sample comprised the 58 worst-performing public schools in the region, whereas our sample was drawn from the entire population of 992 grade 1–9 schools in Rio de Janeiro. Finally, our results differed significantly. While our treatment substantially impacted student learning in mathematics and reading, the experiment conducted by Fryer (2017a) had a near-zero impact when subjects were analysed separately by reading, science, and mathematics.

Our study contributes to these mentioned not-so-clear positive findings with strong evidence of the positive effects of management on learning by accurately assessing management variation in each school, unlike previous studies, and measuring not only the

causal link between the intervention and student learning as well as that between variations in specific management practices due to the treatment and student educational outcomes. Details about the discussion on the previous unclear positive literature are in Appendix A.

Our experiment also contributes to the education literature. A review of 150 education impact evaluations in low- and middle-income countries showed that half of the interventions did not affect learning (Angrist et al. 2021). An extensive review of 747 experiments (1,942 effect sizes) on interventions in education conducted by Kraft (2020) considered that large experiments ($>2,000$ students) with impact estimates on learning greater than 0.1 SD should be considered impressive. The large-scale RCT conducted with 31,760 children reported in this paper shows student achievement effects estimates more than twice that threshold.

Our cost-effectiveness analysis, which follows the J-PAL approach discussed by Dhaliwal et al. (2013), shows that, compared with 27 other relevant interventions in developing countries aimed at improving student learning, the SMEP is one of the most cost-effective at 1.71 “Additional SD per \$100 (PPP)” taking into account the ATE estimates and 4.80 “Additional SD per \$100 (PPP)” considering IV estimates (high-implementation schools). Moreover, our cost-benefit analysis shows that the SMEP has the highest internal rates of return (IRR), 165% for ATE estimates and 227% for IV estimates, compared to 17 relevant education interventions in the United States (Krueger 2003; Fryer 2017a). Since most interventions are ineffective in improving learning, and those that are effective tend to be expensive (Fryer 2017a; J-PAL 2020; Kraft 2020; Angrist et al. 2025), this study contributes to the policy debate by providing a real-world programme that can boost student learning at a very low cost.

II. BACKGROUND AND PROGRAMME DETAILS

II.A. Institutional Background - Rio de Janeiro Education

The municipality of Rio de Janeiro has an approximate population of 6.2 million people and is the capital city of the Federated State of Rio de Janeiro (RJ) in Brazil.⁴ The municipality has one of the largest educational systems in Latin America, responsible for 1,549 educational units and 615,338 students in 2023.⁵ This large state-run system of schools and students is managed by the Municipal Secretariat of Rio de Janeiro (SMERJ). In this study, we use public and state-run schools interchangeably.

Most Rio de Janeiro public schools offer some combination of grades 1-9 (991 schools and 441,842 students in 2023). Other units are nurseries and schools that offer special and adult education. Other units include nurseries and schools offering special and adult education. After grade 9, students are no longer in the Rio de Janeiro education system. The Federated State of Rio de Janeiro provides high school (upper secondary) education.

Public school principals in Rio de Janeiro are elected for two-year terms, with staff, students, and parents eligible to vote.⁶ Only teachers who are city civil servants with at least five years of teaching experience are eligible to become school principals. Principals are generally re-elected several times for the same school. Approximately 85% of principals who served in the 2020–2021 term were re-elected for the 2022–2023 term. Principals are elected with a deputy principal and can choose their staff, including a pedagogical coordinator.

Table A1 in Appendix C shows that the average public expenditure per student by the Rio de Janeiro municipality increased by 102% from 2007 to 2019, whereas that for all Brazilian municipalities increased by 63%. According to the 2020 Court of Accounts of Rio de Janeiro (TCMRio) Special Report,⁷ the average public school teacher’s monthly wage in Rio de Janeiro in 2007 was R\$ 3,741.00 (values updated to 2019), which reached R\$ 6,300.00 in 2019. However, Rio de Janeiro’s school performance, as mea-

⁴<https://www.ibge.gov.br/cidades-e-estados/rj/rio-de-janeiro.html>

⁵<https://educacao.prefeitura.rio/educacao-em-numeros/>

⁶Rio de Janeiro City’s Law n.504/1984.

⁷See TCMRio (2020).

sured by Brazil’s Basic Education Development Index (Ideb)⁸, has declined significantly compared with other Brazilian municipalities. In the 5th-grade Ideb assessment ranking, Rio de Janeiro fell from 1423rd among all Brazilian municipalities in 2007 to 2529th in 2019. Meanwhile, in the 9th-grade Ideb assessment ranking, Rio de Janeiro dropped 586 positions from 2007 to 2019.

Notably, the increase in public expenditure per student in Brazilian municipalities contrasts with the unchanged performance in mathematics, reading, and sciences by Brazil’s 15-year-old students in the OECD’s Programme for International Student Assessment (PISA),⁹ at least since 2009. Brazil’s performance in PISA is discussed in more detail in Appendix B.

II.B. Science and Management for Education Programme

We designed a new public policy to implement management practices in public schools related to improving student learning. Policy-wise, we aimed to implement a programme to enhance students’ educational outcomes through a low-cost intervention that can be easily adapted to different realities and expanded to large school networks. On the research side, our goal was to investigate the causal effect of school management on student learning. Therefore, the management intervention was developed to be the treatment arm of a large-scale randomised field experiment. We have named it the Science and Management for Education Programme (SMEP).

For this purpose, we convened several meetings with the TCMRio and the SMERJ to discuss the terms and conditions for conducting a randomised field experiment in state-run schools overseen by the SMERJ. Brazilian Courts of Accounts are constitutionally obliged to analyse the efficiency and effectiveness of public policies. A partnership agreement was signed in June 2021 between the SMERJ and TCMRio to make the experiment’s implementation possible and establish the experiment’s confidentiality. Additionally, civil

⁸In 2007, Ideb was created to measure the quality of education. It considers two factors: students’ approval rate and average performance in reading and mathematics assessments (Saeb). The Ideb is calculated separately for 5th and 9th-grade students.

⁹PISA evaluates the aptitude of 15-year-old students to use their knowledge and skills in reading, mathematics, and science to tackle real-life problems.

servants from the SMERJ and TCMRio were allocated to form the implementation team, which had the main task of implementing the programme in schools.

In September 2021, we randomly selected 80 schools from the Rio de Janeiro population of 992 public schools for grades 1–9¹⁰. From this random sample, 40 schools were randomly assigned to treatment and 40 to control using a block pair matching procedure, as detailed in Section III.A. The control schools received no support.

The SMEP provided specific management practices to school managers during day-to-day operations through one-to-one coaching sessions and on-the-job training. School managers included principals, deputy principals, and pedagogical coordinators (or supervisors). School managers included principals, deputy principals, and pedagogical coordinators (or supervisors). The intervention aimed to provide schools with the 23 “best” school management practices, as described by Bloom et al. (2015). These practices comprise five groups: operations, monitoring, target setting, people management, and leadership. See Appendix D for more details of these practices.

The implementation was based on the principles of Agile Project Management methodology. This approach enables the incremental and interactive implementation of the SMEP, delivering “small packages” of the policy at each iteration (sprint) (Project Management Institute 2017, 2021).

These school management practices are similar to those outlined by the WMS in different sectors, such as manufacturing, retail, and healthcare, with adjustments to fit the school context (Bloom and Van Reenen 2007, 2010; Bloom et al. 2014, 2015). For each of the 23 ‘best’ management practices, the WMS defines what would be the worst (score 1), the mid-way (score 3) and the best management scenario (score 5) in a school.

Civil servants have much more in common with public school managers than private consulting teams or researchers because they share similar knowledge and experiences regarding Rio de Janeiro’s public administration. Therefore, the design of the implementation team (consisting of civil servants) helped address the challenging issue of gaining the trust of the public school managers.

¹⁰Educational units exclusively responsible for special or adult education and nurseries were excluded from the population of schools.

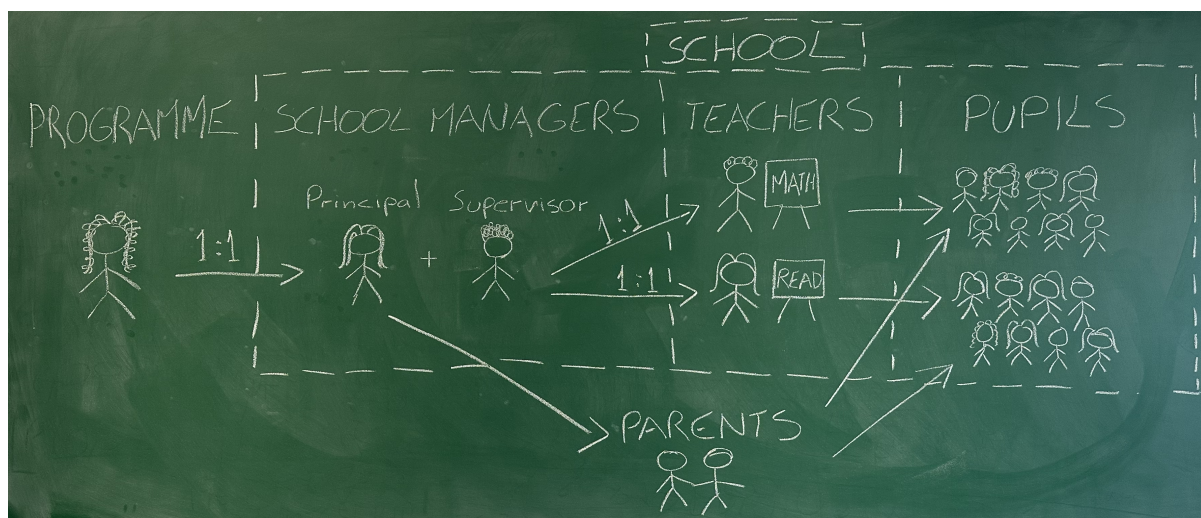
Although the team planned and reviewed the actions jointly in each iteration, the programme was implemented individually or in pairs. Each team member was responsible for the same school for at least six months to ensure some level of trust between the team members and the school managers. Mutual trust between team members and school managers can facilitate the adoption of the best management practices. However, close relationships between team members and school managers may prevent the goal of maintaining some discomfort among managers. The discomfort is supposedly needed to encourage behaviour change. Therefore, we limited each member’s responsibility for the same school to, at most, 1 year.

Through one-on-one coaching sessions with principals (and deputy principals) and pedagogical coordinators (and supervisors), the team discussed the needs, goals, and possibilities for improving the school based on the 23 best management practices. When the team noticed that a school manager could not discuss or implement certain practices due to a lack of skill or knowledge of managerial practices, they provided on-the-job training to address such specific issues. On-the-job training was conducted through one-to-one sessions at the school or through practical thematic seminars.

As the intervention was implemented through one-to-one coaching sessions and on-the-job training, there was no contact between the implementation team and teachers or students. Although participation in the programme was not compulsory, the implementation team could contact the schools whenever necessary to influence them to increase participation. For example, even more resistant schools received team visits for two years to discuss management practices. The degree of adherence to the programme was heterogeneous, as discussed in Section III.

Principals, deputy principals, and pedagogical coordinators (or supervisors) oversee the school’s learning outcomes. However, they do not directly provide educational services to students, which is the role of teachers. The SMEP supported school managers in conducting one-to-one meetings with their teachers, following Bambrick-Santoyo (2018). These meetings were the primary channel for implementing the 23 “best” managerial practices in schools. The SMEP required that school managers hold one-on-one meetings

Figure I
Programme Flow



Notes: The figure illustrates the causal way from the intervention to students' learning. A member of the programme implementation team conducts one-to-one coaching sessions and on-the-job training with school managers from treatment schools to discuss the management of the schools regarding the 23 best managerial practices. School managers from treatment schools, equipped with better management skills, conduct one-to-one sessions with teachers to implement the best management practices. Better-managed teachers provide better teaching to students. Team members also trained the treatment principals to engage parents in their children's education, for example, providing relevant information about the returns to education to their children.

with each teacher at least once every two months.

Figure I illustrates the causal path of our intervention in student learning. A member of the SMEP team conducts one-to-one coaching sessions and on-the-job training with school managers in treatment schools to discuss the 23 best practices. With improved management skills, school managers hold one-to-one sessions with teachers to facilitate the implementation of most of these practices. Consequently, the teachers are able to deliver more effective instruction to students. Principals also receive training to engage parents in their children's education by, for example, providing relevant information about the returns of education.

One final point concerns the challenge of accessing schools in areas of Rio de Janeiro controlled by criminal gangs, where even police face access difficulties. Despite efforts to reach these schools, it was impossible to access some, as in the case of an armed conflict near the school. When schools were inaccessible due to violence and crime, the managers of these schools were brought to the implementation team's office. The school

managers reported feeling strongly valued by such an approach. This challenge highlights the heterogeneous school realities across Rio de Janeiro. The following section outlines the SMEP’s implementation timeline from January 2022 to December 2023.

II.C. Programme Implementation - 2022 to 2023

The SMEP worked through iterations (sprints). The first sprint lasted two months, between January and February 2022. It consisted of an opening seminar, one-to-one coaching sessions, and on-the-job training on two related management practices: “Standardisation of Instructional Processes” and “Data-Driven Planning and Student Transitions.”

The opening seminar introduced the programme and encouraged the school managers to participate in the initiative (participation was not compulsory). The team provided schools with feedback sessions on their management level as per a survey conducted at the end of 2021 (before the random assignment) to diagnose school management levels. This survey assessed the management of schools on the 23 best management practices. More details on the survey conducted in 2021 are in Section III.B.

In this sprint, the managers were clearly resistant to using student data for direct actions. The team had to work carefully to overcome this resistance. Furthermore, there was a lack of standard data-analysis skills. For example, most school managers were unable to work with spreadsheets. Consequently, the programme provided more intense on-the-job training sessions to help managers organise and analyse student data.

The second sprint, which started in March 2022, aimed at providing schools guidance related to the following management-related practices: “Target Balance,” “Target Interconnection,” “Time Horizon of Targets,” “Target Stretch,” and “Clarity and Comparability of Targets.” The programme team chose to work with these practices to help school managers develop action plans demanded by the Secretariat of Education (SMERJ). As the team employed the Agile Management Project methodology, adapting to the demands of the Education Secretary (SMERJ) was simple.

The team used the SMART (specific, measurable, achievable, realistic, and time-

bound) methodology to set goals and targets (Doran et al. 1981). The aim was to help schools set goals and targets and break them down at the teacher and class level. As a part of the process, the team conducted a practical seminar to help principals, principal deputies, and pedagogical coordinators resolve common doubts regarding these management practices.

The third sprint took place in May 2022, and it focused on reviewing previous practices as the implementation team realised that some school managers had not fully absorbed the on-the-job training and were not confident in their ability to conduct one-to-one meetings with teachers. The team continued to assist the managers in developing their school action plans, as required by the SMERJ. At this point, most, but not all, schools in the treatment group were already engaged with the programme and even sought support from the implementation team.

At the end of May, the team started a new sprint focusing on all management practices from the monitoring group: “Continuous Improvement,” “Performance Tracking,” “Performance Review,” “Performance Dialogue,” and “Consequence Management.” This sprint started with a workshop for pedagogical coordinators (supervisors), focusing on the monitoring group of management practices. The team again highlighted the relevance of one-to-one coaching sessions with teachers to improve school management.

After the mid-year school holidays in Brazil (July), the team continued developing the monitoring and target-setting management practices (August 2022). The team also started discussing “The Personalization of Instruction and Learning Practice.” This sprint lasted 1 month.

In September and October 2022, the team changed the focus to leadership practices. Principals, deputy principals, and pedagogical coordinators were invited to participate in a full-day practical workshop on leadership. They had to solve school practical problems related to three management practices: “Leadership Vision,” “Clearly Defined Accountability for School Leaders,” and “Clearly Defined Leadership and Teacher Roles.” During the workshop, school managers were encouraged to consider the management practice of “Adopting Educational Best Practices” from the operating group of the 23 best manage-

ment practices.

Following the workshop, this two-month sprint included one-on-one sessions with school managers to deepen their understanding of leadership practices and “Adopting Educational Best Practices.” The team provided school managers with simple slides regarding the returns of education. School managers were encouraged to present these slides to parents and students to increase their awareness of the importance of education. The school managers were informed that providing such information to parents and children could improve students’ frequency of attendance and performance (Nguyen 2008). The school managers provided very positive feedback on the slides, mentioning that parents were interested in the slides; most asked to receive a printed copy.

During the November and December 2022 sprint, the team coached schools on all the aspects discussed throughout the year. The team also discussed ways to help school managers implement the following people management practices: “Rewarding High Performers,” “Removing Poor Performers,” “Promoting High Performers,” “Managing Talent,” “Retaining Talent,” and “Attracting Talent.” However, there was little space for developing these practices. In Brazil, legislation makes it challenging to promote, reward, remove, or fire civil servants, such as public school teachers, based on performance.

The first year of the programme implementation ended in December 2022, with the year’s last seminar coinciding with the end of the school year in Brazil. The team invited school managers who best implemented the program, some from very challenging local settings, to present their experiences to other treatment schools. The goal was also to influence less-engaged schools, showing the success of school managers who had worked in challenging situations.

This event captured the school managers’ attention and motivated them to participate more actively in the SMEP. A post-event survey showed that almost all school managers were satisfied with the seminar and the programme. As local settings vary considerably across Rio de Janeiro, school managers were pleased to see a way to improve school management practices even in the most challenging situations.

SMEP resumed in January 2023, focusing on deepening the school’s adherence to the

23 best management practices. In the first year, school managers were extensively exposed to the best management practices and necessary tools; in contrast, the second year focused on deepening the implementation of these practices. During the first year, on-the-job training was more prevalent than one-to-one coaching sessions; however, the second year primarily relied on the latter. The team continued to provide on-the-job training when necessary, but the focus was mainly on coaching sessions with school managers. With the skills acquired from the first year, school managers developed a less data-resistant mindset, and the coaching meetings with the programme team improved significantly.

In addition to the sprints to coach school managers on the best management practices in 2022, two relevant seminars were conducted in 2023. Both utilized the same dynamic from the last seminar of 2022, with the most engaged school managers, mainly from schools in highly challenging situations, presenting their experiences in advancing management in their schools. The post-event surveys indicated that school managers were very satisfied with their shared experiences. Figure II shows five photographs of the SMEP in action.

III. METHODS AND DATA

III.A. The Selection of Schools for the Management Experiment

This study complies with all critical research protocols. It received approval from the University of Cambridge’s Ethics Committee and is registered on the AEA RCT registry (AEARCTR-0007669 on 15 May 2021; <https://www.socialscisearch.org/trials/7669>). The intervention was not sensitive. Rio de Janeiro authorities and institutions responsible for education in the city also approved the experiment.

We conducted a clustered pair-matching randomised field experiment with 80 schools from the municipality of Rio de Janeiro to analyse the impact of our management programme on students’ educational outcomes. This experiment is clustered because we randomly assigned schools, not pupils, to treatment. It is pair-matching because we formed school pairs based on previous educational achievements, and within each pair, we randomly assigned one school to treatment and the other to control. It is considered a

Figure II

Science and Management for Education Programme (SMEP)



Notes: The top left photo shows a principal organising the students' educational performance data for one-on-one coaching sessions with teachers. The top-right photo depicts an implementer coaching school managers on the target-setting group of management practices. The central picture depicts the TCMRio auditorium with school managers after a day of hands-on leadership training. The bottom-left photo shows a principal presenting information provided by the program regarding the returns of education to parents. The bottom-right image highlights two implementation team members coaching a principal, observed by the pedagogical coordinator.

field experiment because it was conducted as an ordinary pilot government project within the Brazilian public administration.

The sample of 80 schools was randomly sampled from the 2021 Rio de Janeiro population of 992 grade 1–9 public schools. These schools are of two main types: first-segment schools,¹¹ responsible for grades 1–5¹², and second-segment schools,¹³ responsible for grades 6–9. First-segment schools represent 70.9% and second-segment schools 29.1% of schools under the SMERJ.

The participant schools were unaware that they were part of an experiment, and the sample identification was kept confidential. The participating organisations signed a confidentiality agreement to protect the experiment’s integrity. Only one implementation team member had access to the sample identification and was required to keep the information strictly confidential. The term “experiment” was avoided in the programme implementation vocabulary. The project was presented to the schools as a typical initiative conducted in public administration. These measures were implemented to minimise the risk of bias and spillover.

One of the differences between first- and second-segment schools is their teaching approach. First-segment schools have generalist teachers who teach multiple subjects to one class, whereas second-segment schools have specialist teachers who teach one subject to various classes. Second-segment school managers have more teachers to manage, making their jobs more complex. Our random sample of 80 schools had 56 schools from segment I (grades 1–5) and 24 from segment II (grades 6–9). Thus, we conducted separate random assignments for first- and second-segment schools to ensure equal representation in both groups. This method allowed us to assess the causal effect of our programme across and within both segments. Block randomisation may also generate smaller standard errors (SEs) without a downside for the analysis (Gerber 2012).

Using information correlated to the outcome of interest in the random assignment can avoid the emergence of large and significant random differences in the outcomes between

¹¹Also known as primary schools in the UK and the US.

¹²As a rule, first-segment schools include grades 1–5, although some also include grade 6.

¹³Also known as lower secondary schools in the UK and the US

treatment and control groups before the experiment begins (Abadie and Imbens 2011; Fryer 2017a). One of our main outcomes of interest is student learning. At the time of the assignment, we could not access the scores from the standardised reading and mathematics tests, the Rio Assessments, applied by the SMERJ to all students in 2021. The only student learning assessment results available in October 2021 were from the 2019 Brazilian Assessment System for Basic Education, or Saeb 2019. The Saeb 2019 comprises mathematics and reading tests intended to be applied nationally to all 5th- and 9th-grade students.¹⁴

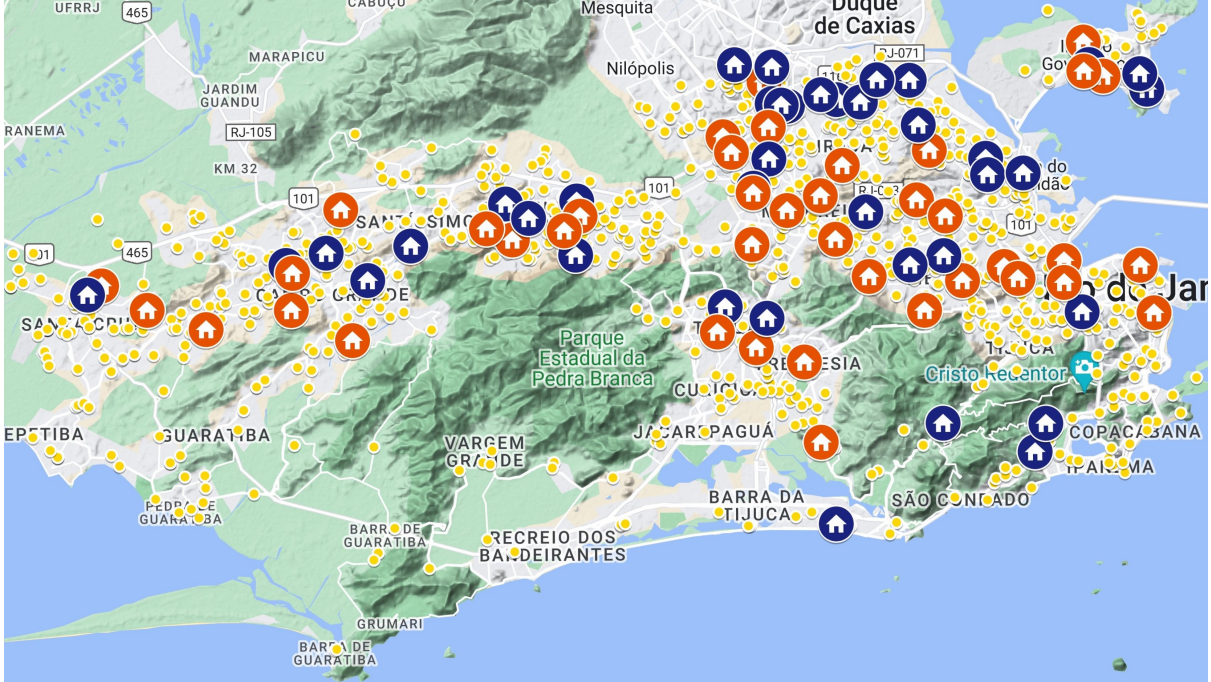
We adopted a pair-matching procedure to perform the random assignment within each school block (segment). First, we created a school ranking by block, first- and second-segment schools, based on the descending order of the average school score in the Saeb 2019 (sum of reading and mathematics school averages divided by two). First-segment schools were represented by the 5th-grade average score, whereas second-segment schools were represented by the 9th-grade average score.

Within each block (school segment), we arranged pairs of schools in descending order per the Saeb 2019 average and randomly assigned one school each to the treatment and the control groups. This process was repeated for each block until all 40 pairs were assigned. We generated four groups of schools: treatment and control groups for first-segment schools (28 schools each) and for second-segment schools (12 schools each). This method ensured balance in terms of school segment and Saeb 2019 school average scores across the treatment and control groups.

Figure III presents a map of the municipality of Rio de Janeiro with the geographical distribution of the 992 grade 1–9 state-run schools in 2021. Blue and red circles represent the treatment and control schools, respectively, and yellow dots represent the remaining schools. The distribution of the treatment and control groups followed the population distribution pattern. The next section details the process for generating data on the schools’ management level before (September 2021) and after (December 2023) the intervention.

¹⁴Only scores from schools where more than 80% of their 5th/9th-grade students participated in the 2019 Saeb were accessible.

Figure III
Geographical View of Schools in Rio de Janeiro



Notes: Blue and red circles denote treatment and control schools, respectively. Yellow dots represent the remaining grade 1–9 public schools in Rio de Janeiro. *Source:* Created using My Map from Google Maps. The SMERJ provided the schools’ address information.

During October and November 2021, principal candidates participated in an intense certification course necessary to be eligible for the 2022–2023 term election.¹⁵ Generally, an “election” has only one candidate for each school. Thus, when the random assignment was conducted in October 2021, each school’s principal for the 2022–2023 term was already defined. According to the SMERJ, no sample schools’ principal candidates changed schools for the 2022–2023 term, abandoned the certification course needed to be a principal candidate, or withdrew their candidacy for the 2022–2023 term election.

School principals are commonly elected multiple times for the same school in Rio de Janeiro. Data on the number of schools with the same principal in the 2020–2021 and 2022–2023 terms helps clarify the point. For the 2022–2023 term, 82.5% and 87.5% of principals remained the same from the 2020–2021 term for the treatment and control groups, respectively. We define a dummy that takes a value of 1 when the principal did not change between the 2020–2021 and 2022–2023 terms and 0 otherwise. We regress

¹⁵The documentation regarding the Rio de Janeiro school principals’ election for the 2022–2023 term can be found at https://doweb.rio.rj.gov.br/apifront/portal/edicoes/imprimir_materia/761940/5099.

this dummy variable on a treatment variable using robust SEs. The regression shows a difference of 0.05 with a SE of 0.08. That is, the difference is not statistically significant.

It is also improbable that teachers have left schools or self-selected for treatment in 2021 based on the treatment assignment, as teachers are mainly chosen to work in a school by the SMERJ based on school needs. For the same reason that teachers cannot choose the schools they want to work in, it is very unlikely that teachers have left schools or self-selected for treatment during the programme implementation.

It is important to mention that, due to the treatment assignment information, it is unlikely that either students or their parents self-selected for treatment in 2021. A strict confidentiality agreement was in place to protect the experiment. Furthermore, the programme did not work directly with students. Instead, it focused on school management, specifically principals, deputy principals, and pedagogical coordinators.

As only the researcher knew which schools were the controls and no actions were directed at them, any discussion of bias in the control group was unnecessary.

III.B. Data I - Management Level of Schools in 2021 and 2023

We conducted two surveys with the 80 selected sample schools. The first was conducted just before the random assignment in September 2021, and the second was conducted in December 2023, after the intervention. The surveys were double-blind, as school managers were unaware that their schools were being scored, and interviewers were unaware of the schools' educational performance. We closely followed the survey methodology developed by the WMS (Bloom and Van Reenen 2007, 2010; Bloom et al. 2014, 2015). The only difference between our survey and the WMS methodology is that we conducted in-person interviews instead of by telephone.

The 2021 survey aimed to diagnose the management level of the schools in our sample and train the implementation team to deliver the treatment. The 2023 survey had two main goals: (i) to analyse the efficacy of the treatment in the implementation of the 23 best management practices, which is the same as assessing the treatment's impact on the school's management, and (ii) to analyse the causal effect of school management practices

on student learning.

We used the WMS questionnaire and scoring grid developed by Bloom et al. (2015) to measure the management of the sample schools in terms of the 23 best management practices.¹⁶ The questionnaire contains 23 dimensions, each corresponding to one of the 23 management practices. Based on deep interviews, each management practice was scored against a grading score ranging from 1, the worst, to 5, the best management level.

For instance, in practice 1) of the questionnaire, “Standardisation of the Instructional Processes,” a school receives a score of 1 if “no clear or institutionalised instructional planning processes or protocols exist; little verification or follow-up is done to ensure consistency across classrooms”; a score of 3 if the “school has defined instructional planning processes or protocols to support instructional strategies and materials and incorporate some flexibility to meet students’ needs; monitoring is only adequate”; and a score of 5 if the “school has implemented a clearly defined instructional planning process designed to align instructional strategies and materials with learning expectations and incorporate flexibility to meet student needs; these are followed up on through comprehensive monitoring or oversight.”

The WMS questionnaire comprises open-ended questions that avoid leading respondents toward a particular answer. For example, the initial inquiry in the performance monitoring aspect asks, “What kind of main indicators do you use to track school performance?” rather than a close-ended query such as “Do you evaluate your school’s performance using the academic scores of students [yes/no]?” (Bloom et al. 2015). The initial open-ended inquiry is succeeded by additional questions, such as “How frequently are these indicators measured?,” “Who gets to see this data?,” and “If I were to walk through your school, what could I tell about how you are doing against your indicators?” The interviewers could ask follow-up questions when necessary. The goal was to give scores regarding the actual management practices of the school and not the school managers’ aspirations of what the school’s management should be.

¹⁶The questionnaire and scoring grid can be found at <https://worldmanagementsurvey.org/data/dwms-public-sector/questionnaires/>.

The WMS also developed an updated and more detailed version of the scoring grid contained in the original version of the questionnaire, called D-WMS.¹⁷ Both scoring grids, the original WMS and D-WMS, are compatible, ranging from 1, worst, to 5, best. We tested the D-WMS scoring grid in some schools outside our experimental sample. The D-WMS tool was much more time-consuming and challenging regarding the interviewers' training. Scoring schools was also more time-consuming and difficult as the interviewers needed to learn and think about many more scoring possibilities than the original version. Therefore, we opted to use the original WMS questionnaire.

The implementation team conducted the management survey in September 2021. As the random treatment assignment had not yet occurred, there were no concerns about bias from the team at the time. However, after two years of intervention, the team was deeply involved with the treatment schools, which could be a clear source of bias in the second survey. Therefore, we requested 1 of our partners, the TCMRio, that seven auditors not involved with the experiment would conduct the December 2023 survey. The Court agreed to allocate these auditors from November to December 2023 for training sessions and conducting the 2023 management survey.

The interviewers in 2021 and 2023 had an intensive two-week training on the WMS methodology. The survey teams understood how Rio de Janeiro state-run schools functioned as they were bureaucrats who specialised in education and public management from the municipal public administration. This knowledge was crucial to ensure that the practices observed in schools were adequately codified as proposed by Bloom and Van Reenen 2007, 2010; Bloom et al. 2014, 2015. As part of the training sessions, the teams gained practical experience applying the WMS survey tool to some schools selected from outside the study sample. After the training surveys, the interviewers discussed the scoring procedure. The goal was to align the scoring patterns across the members of the survey teams.

The SMERJ contacted the 80 sample schools for both surveys (2021 and 2023) and informed them that interviews would be conducted with principals to discuss local school

¹⁷<https://worldmanagementsurvey.org/data/dwms-public-sector/methodology/>.

operations. No additional information was provided. School principals have considerable knowledge of the management level of their schools and are involved in day-to-day school activities. Other senior managers, such as the secretary, superintendent, and region coordinator, supposedly work with more strategic actions far from regular school activities. The SMERJ was informed not to disclose the nature of the survey. The word “interview” was adopted to obfuscate the idea that participant schools would be graded, following Bloom and Van Reenen 2007.

We applied the procedures recommended by the WMS to avoid potential survey bias (Bloom and Van Reenen 2007; Bloom et al. 2015). Each team member received a list with only the names and contact details of the schools. The schools were distributed randomly to interviewers, but no interviewers could apply the survey to any schools in their neighbourhoods. This ensured that schools were not known to interviewers. Additionally, the interviewers were instructed not to obtain more information about their allocated schools. They were asked to contact their allocated schools and arrange face-to-face interviews with principals. They were also advised not to discuss students’ educational results.

One relevant point stressed by the WMS methodology is that the survey should focus on managerial practices, not the principals’ views on how schools should be managed. The interviewers were allowed to make annotations; however, they were not to score the school in front of the principal. The coding process took place after the interviews and outside school facilities. The interviews were recorded, allowing the interviewers to consult the recording. The interviews in both surveys lasted, on average, 75 minutes.

We achieved a 100% response rate in both surveys; that is, all the principals of the 80 schools participated in the two surveys. This is a much higher response rate than in other studies conducted across different sectors by the WMS, which have recorded an average 41% survey response rate, according to Bloom et al. 2015.

We double scored all interviews in both surveys, such that each management practice in each school received two different scores from two different coders (also called raters). The first interviewer (coder or rater) gave each school’s first (primary) scores. Meanwhile, the second scoring round for each school was conducted using interview recordings by

randomly chosen members. The goal was to obtain inter-rater reliability or agreement coefficients.

The management level of a school should be an attribute of the school and not of any rater (Gwet 2021). Appendix E shows high agreement coefficients in both surveys, indicating that the raters' scores assigned for each of the 23 management practices for each school were independent of the rater who evaluated the school. Table A3 in Online Appendix E reports a Brennan–Prediger coefficient of 0.720 and a Gwet's AC2 of 0.760 for the 2021 survey. The 2023 survey presents higher agreement coefficients, with a Brennan–Prediger coefficient of 0.788 and a Gwet's AC2 of 0.811. These values are considered by Landis and Koch (1977) to be relatively high or close to perfect agreement between raters. This evidence suggests that the management scores given to schools did not reflect the raters' subjective views but the schools' actual characteristics.

The scores represent the schools' management practices in three different formats. First, the general management of a school is the average of the scores reached by a school regarding each of the 23 management practices surveyed. Second, as the 23 practices can be grouped into five groups—target setting, leadership, operation, monitoring and people management—each group in each school is represented by the average of the scores reached by that school for the practices that form the specific group. As each management practice is scored against a scale ranging from 1 to 5, the general management and group of practice averages also follow the same scale. Third, analyses are also conducted separately with each of the 23 management practices.

III.C. Data II - Other Variables

The SMERJ provided the following pretreatment data at the individual level: student race, gender, grade (1–8), 2021 Rio Assessments scores, absence rate, and whether the student's family in 2021 was a beneficiary of the Brazilian cash transfer programme (Bolsa Família); race, gender, age, education, and experience of principals and pedagogical coordinators; and teacher gender and education. At the school level, the SMERJ provided information on school size by number of students and school segment (segment I, 1st to

5th grade; segment II, 6th to 9th grade). We also generated data on the management level of schools in 2021 through a management survey conducted in the same year. All these covariates were collected before January 2022 and considered pretreatment covariates. Regarding the 2021 Rio Assessments, the scores were based on the classical test theory (CTT; rate of correct answers).

Unfortunately, approximately $\frac{1}{3}$ of students' scores from the Rio Assessments in 2021 were missing. Many students were not assessed due to the COVID-19 pandemic. Therefore, to address missing data in the 2021 scores as well as other pretreatment covariates, we used the missing indicator method (Kayembe et al. 2022; Zhao and Ding 2022). Another issue with the 2021 Rio Assessments was where the students took the tests. Although the Rio Assessments were designed to be taken at school, they were also allowed to be taken from home due to the COVID-19 pandemic, making it difficult to control for interference.

The primary outcome variable comes from the Rio Assessments, which are administered four times each academic year—in April, June, September, and December—to all first- to ninth-grade students.¹⁸ These tests were developed by a renowned centre of studies in education assessments from a federal university in Brazil. The centre also coordinated and supervised test applications. The tests began to be administered in Rio de Janeiro in April 2021. Further information on the application of the tests is in Online Appendix F.

The Rio Assessments comprise reading and mathematics tests based on the Common National Curriculum Base (BNCC).¹⁹ Our primary analysis relies on the Rio Assessments scores that use the item response theory (IRT) to score students against the same cross-grade scale of competencies as the Saeb.²⁰ Thus, students from different grades are scored based on the same competency scale, allowing comparability across grades.

Based on the December 2023 Rio Assessments and considering only control students, we observed that the average difference in students' scores by grade was 13.14 score points

¹⁸In Portuguese, this is called *Atividade Diagnóstica em Rede (ADR)*.

¹⁹See <http://basenacionalcomum.mec.gov.br>.

²⁰For Saeb scales, see the webpage (<https://www.gov.br/inep/pt-br>) of the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inpe), the agency connected to the Brazilian Ministry of Education in charge of evaluating educational systems and the quality of education in the country.

for reading and 14.73 for mathematics.²¹

Figure A5 in Appendix F shows the control school students' scores by grade for reading and mathematics in the December 2023 Rio Assessments. Boxplots represent the distribution of scores for each grade. The average score from the end of the 2nd to the end of the 9th grade varied by around 100 points in reading and mathematics. With seven years of schooling, students acquired, on average, 100 score points in reading and mathematics in Rio de Janeiro. However, there is significant variance across grades. The greatest reading and mathematics learning gains occurred through the 4th and 5th grades.

Table A4 in Appendix F provides data on students' missing reading and mathematics scores between 2022 and 2023. The fixed sample data comprised 31,760 students from the 80 sample schools. The first two columns of the table, labelled "Before adjustment," show the number and percentage of missing scores of the complete sample before any changes were made to fill in the missing data. The missing data can be attributed to many reasons, such as students being transferred to schools outside the sample, being unable to attend school on the test day, or moving from a treatment to a control school or vice versa. The fourth and fifth columns, labelled "After adjustment," show the missing data after filling in each student's missing data with their standardised scores from previous tests on the same subject.

For instance, if a student does not have the standardised reading score from the December 2023 Rio Assessments, we use the standardised reading score from the September 2023 Rio Assessments as a replacement. Meanwhile, if the September 2023 Rio Assessments reading score is also missing, we use the standardised score from the June 2023 Rio Assessments. If necessary, we continue this process until April 2022 (the first Rio Assessment after the beginning of the treatment). We only replace missing data with previous standardised scores to maintain a student's relative position in their population grade in a specific subject and Rio Assessments application date. As the scores were stan-

²¹Only students from grades 1 to 8 at the beginning of the 2022 school year were considered in our fixed sample. However, some students failed to go to the next grade in 2023; 25 students in grade 1 in 2022 remained in the same grade in 2023. As only these 25 students represented the 1st grade in 2023 in our sample, we did not use this grade in our analysis.

dardised by subject, grade, and application date using the control group as the reference, replacing missing data with information from previous tests is a conservative approach, as students' older scores are from when they were less exposed to the programme.

Based on this adjustment, our attrition rate is almost null, that is, only 1.10% of students' missing data for reading and 1.12% for mathematics. Regressions of a dummy representing missing data after the adjustment on a treatment dummy show that the attrition rate differences between treatment and control groups are 0.16% for reading and 0.18% for mathematics; neither difference is statistically different from 0 at standard confidence levels. Even before the adjustment, the differences were 0.34% for reading and 0.13% for mathematics, which are not statistically significant.

From SMERJ, we also had access to the scores from the Rio Assessments generated using the Classical Test Theory (CTT), i.e., the scores were the simple proportion of correct answers. Following the same procedure as the IRT scores, we replaced missing data in these CTT scores.

III.D. PreTreatment Summary Statistics

The school year in Brazil runs from January to December, with the regular enrolment period taking place between November and December of the previous year. We focused on students enrolled for the 2022 academic year during the standard enrolment period from November to December 2021. We did not consider students enrolled later in the 2022 academic year. As pretreatment variables, we collected student characteristics at the end of 2021.

As previously mentioned, we excluded students enrolled in the 9th grade for the 2022 academic year as they were supposed to attend upper-secondary schools (high schools) in 2023, which is outside the scope of the study. The treatment was planned for two years, from January 2022 to December 2023. Consequently, the student sample in this study comprises 1st- to 8th-grade students enrolled (during November/December 2021) in a sample school for the 2022 school year. Our sample included 31,760 students from 80 state-run schools. In the 2022 school year, 16,134 and 15,626 grade 1–8 students

were enrolled in the control and treatment schools during the regular enrolment period (November and December 2021).

Table I: Pre-Treatment Statistics - Student Sample

	Valid obs	Control (mean)	Treatment (difference/SE)
A. Characteristics			
Female (%)	31,427	0.488	-0.003 (0.006)
White (%)	28,743	0.335	0.019 (0.019)
Brown (%)	28,743	0.543	-0.015 (0.013)
Black (%)	28,743	0.120	-0.003 (0.010)
Bolsa família (% receiving)	31,427	0.340	-0.060 (0.033)
B. Educational outcomes			
Reading - April 2021	20,110	0.000	0.074 (0.056)
Mathematics - April 2021	19,982	0.000	0.054 (0.058)
Reading - June 2021	21,391	0.000	0.081 (0.058)
Mathematics - June 2021	21,309	0.000	0.066 (0.059)
Students total (Schools)	31,760 (80 schools)	16,134 (40 schools)	15,626 (40 schools)

Notes: This table presents information on the characteristics and educational outcomes of students enrolled (during November/December 2021) in a sample school for the 2022 school year. The column “Valid obs” shows the number of observations with data available. The “Treatment” column shows the difference between the treatment and control averages with the standard errors (SE) of the differences reported in parentheses. “SE” means clustered robust standard errors with clusters defined as the school pairs used for the random assignment. Each difference between averages and standard error comes from a regression analysis where relevant variables in the table are regressed on a treatment dummy. Students’ scores from the 2021 Rio Assessments were standardised by subject, grade, and test application date using the control group as the reference.

According to Table I, Black students comprised 12% of the control group and 11.7% of the treatment group, whereas Brown students accounted for 54.3% and 52.8% in the control and treatment groups, respectively. A total of 34% of the students in the control group belonged to beneficiary families in Brazil’s Bolsa Familia cash transfer programme, whereas this percentage in the treatment group was 28%. The scores were slightly higher in the treatment group regarding student educational outcomes. As the control group was the reference group, the

average score of the control group was 0.²² The differences encountered are due to chance because we randomised the treatment.

Table II: Pre-Treatment Statistics - School Sample

	Valid obs	Control (mean)	Treatment (difference/SE)
Principals			
Female (%)	80	0.900	-0.100 (0.080)
White (%)	69	0.459	-0.053 (0.121)
Brown (%)	69	0.405	0.126 (0.121)
Black (%)	69	0.135	-0.073 (0.072)
College degree (%)	80	0.725	0.025 (0.100)
Experience (years)	70	6.4	1.0 (1.5)
Pedagogical Coordinators			
Female (%)	75	0.973	-0.078 (0.057)
White (%)	58	0.645	-0.127 (0.131)
Brown (%)	58	0.323	0.011 (0.126)
Black (%)	58	0.032	0.116 (0.077)
College degree (%)	74	0.811	0.027 (0.090)
Experience (years)	60	3.1	0.0 (0.7)
Teachers (average by school)			
Female (%)	80	0.814	-0.002 (0.034)
College degree (%)	80	0.816	0.035 (0.032)

Notes: This table presents the characteristics of principals, pedagogical coordinators and teachers in the 80 schools sample for the 2022 school year. The column “Valid obs” shows the number of observations with data available. The “Treatment” column shows the difference between the treatment and control averages with the standard errors (SE) of the differences reported in parentheses. “SE” are clustered SE with the clusters defined as the school pairs used for the random assignment. Each difference between averages and SE comes from a regression analysis where relevant variables in the table are regressed on a treatment dummy.

²²The educational outcomes in Table I are taken from the Rio Assessments, which were conducted before the random assignment on two different dates: April and June 2021.²³ Scores from the 2021 Rio Assessments were standardised by subject, grade, and test application date using the control group as the reference.

According to Table II, most principals and pedagogical coordinators were White females. Additionally, the average experience of principals was 6.4 and 7.4 years for control and treatment schools, respectively, and the average experience for pedagogical coordinators was 3.1 years for both groups. Most principals, pedagogical coordinators, and teachers in both groups had a college degree.

Regarding the management level of the schools before the experiment, a regression of the 2021 school management level on a treatment dummy using robust SE shows a nonsignificant difference of 0.027 (0.857) score points between treatment and control school management levels²⁴.

The 2021 management level of control schools was 2.407, and that of treatment schools was 2.379. This means that the 23 best management practices were implemented at a relatively low level, not achieving even the midway management level score of 3. Figure IV shows that not only were treatment and control schools very similar regarding point estimates but also in terms of their overall management level distributions (1 - CDF).

Figure V reports a similar probability by grouping the 23 management practices into five groups: target setting, leadership, operation, monitoring and people management. Almost 90% of schools did not reach level 2 in the people management group. Moreover, approximately 50% of the schools' levels in target setting and leadership were lower than 2. The schools were relatively better managed in the operation and monitoring groups. As expected from a random assignment procedure, the control and treatment groups reported very similar cumulative probabilities across all five groups.

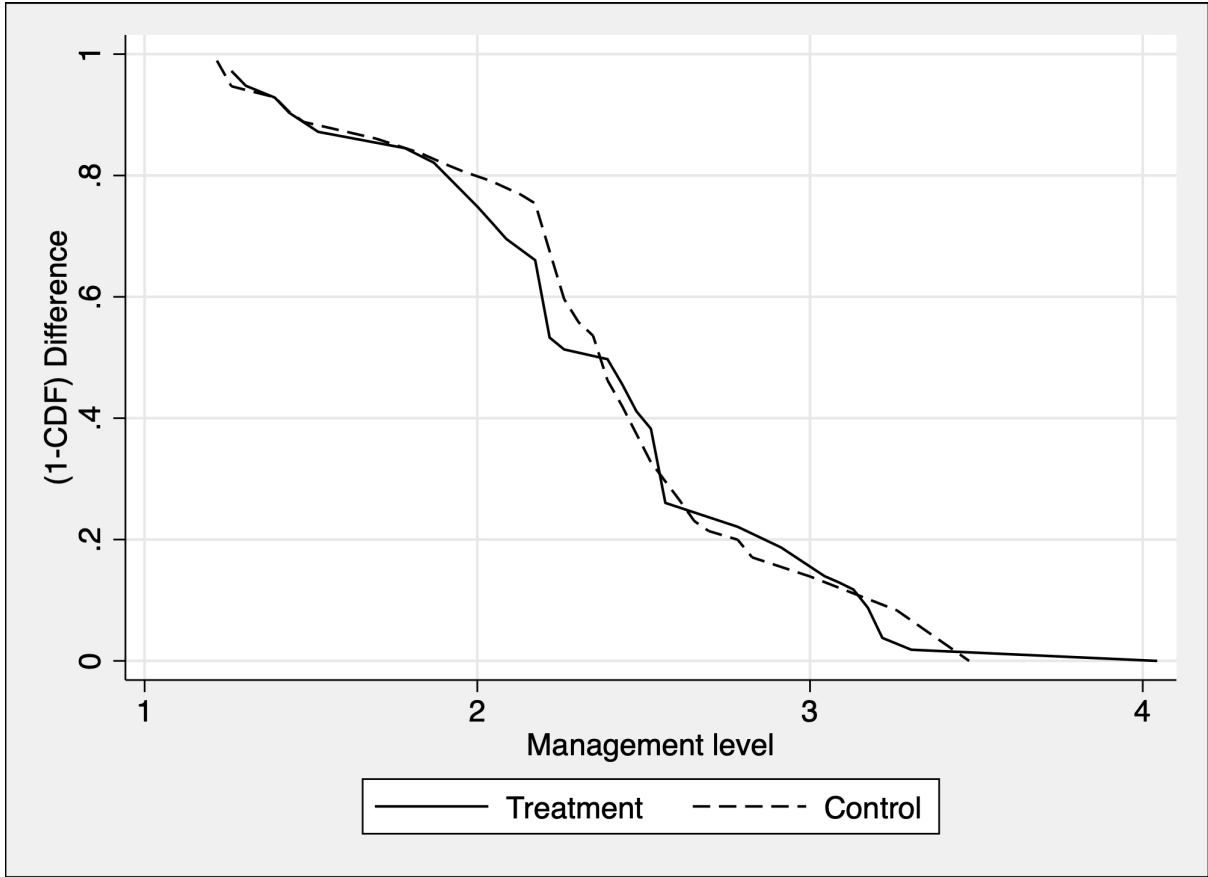
Based on Table I, Table II, Figure IV and Figure V, we conclude that the treatment and control groups were, on average, very similar in 2021. Our experimental sample is also very similar, on average, to the population of students and schools in Rio de Janeiro, as shown in the Appendix G.

The fact that the SMERJ and its schools apply the Rio Assessments to students (despite the supervision of the external centre) may raise questions about school managers' and teachers' external influence on the scores of the treatment schools relative to control schools. However, this is unlikely, given how the exams are administered and the design of the experiment.

The SMERJ and schools applied the tests under the close supervision of the specialised

²⁴If the management is standardised to have a mean of 0 and an SD of 1 regarding the control group, the difference is 0.048 (0.267).

Figure IV: Treatment and Control Schools' Management Level in 2021



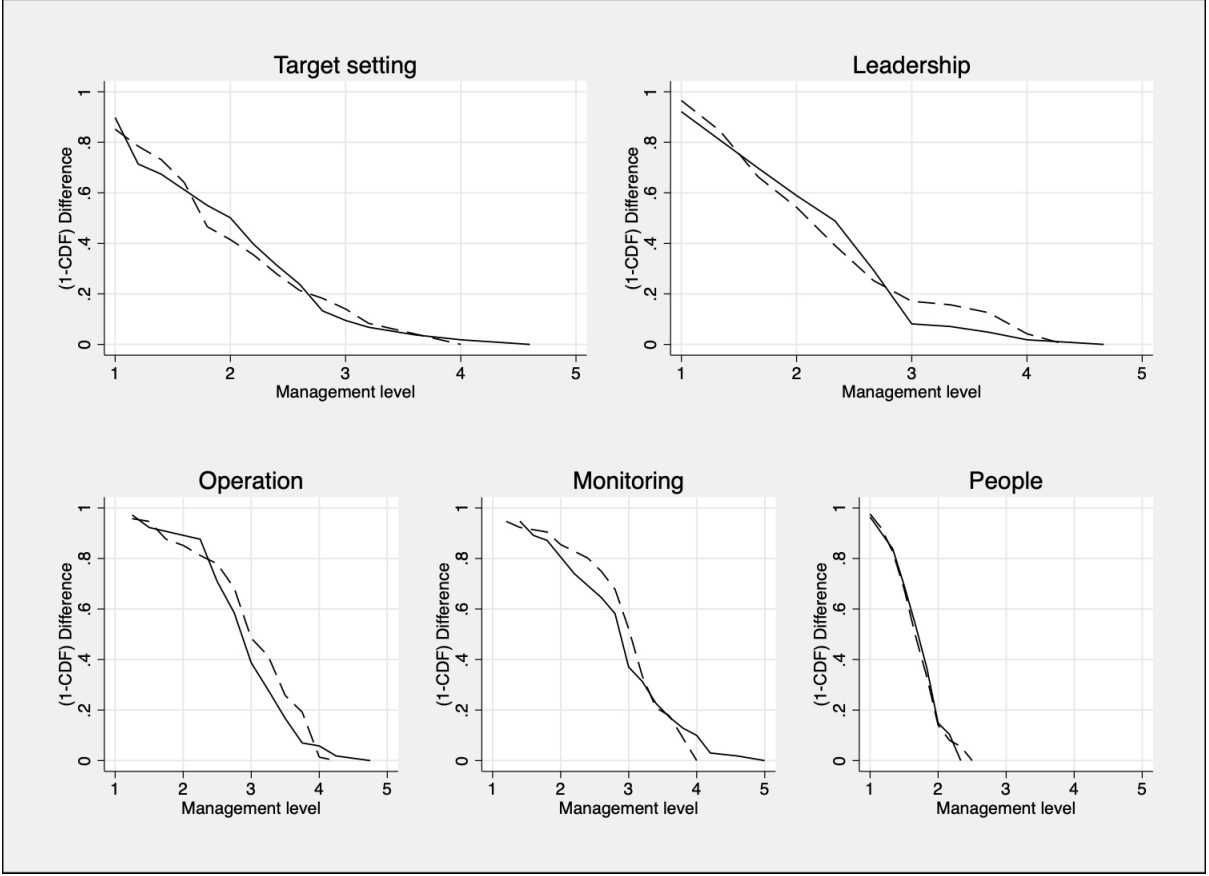
Notes: The figure illustrates the difference between treatment and control schools regarding the probability of a school's management level being greater than a specific management level on the x-axis (i.e., 1 minus the CDF). Although the management level scale ranges from 1 to 5, the x-axis only displays the range where observations are present to improve clarity. The solid and dashed lines represent treatment and control schools, respectively.

centre. Typically, the teachers administering the exams differ from those teaching the classes. Additionally, the answer sheets sent to schools are uniquely identified for each student. Furthermore, as answer sheets contained the identification of each student, it is improbable that school managers could make any changes without the involvement of teachers.

School managers would need to make teachers agree to misbehave to artificially increase their students' scores in the Rio Assessments. Teachers are accountable for administering exams; thus, school managers do not have direct, institutional ways to force teachers to act improperly. They cannot promote, reward, remove, or fire teachers. Teachers have robust protections from Brazil's public administration for resisting illegal or unethical demands. Moreover, the study did not provide additional rewards for principals if school performance improved.

In our experiment design, the treatment and control schools were unaware they were part of a scientific experiment due to the strict confidentiality agreement. Additionally, schools were

Figure V: Treatment and Control Schools by Management Groups in 2021



Notes: This figure illustrates the difference between treatment and control schools regarding the probability that a specific group of management practices at a school are greater or equal to a specific management level on the x-axis (i.e., 1 minus the CDF). The management level on the x-axis ranges from 1 to 5. The solid and dashed lines represent treatment and control schools, respectively.

informed that they were not being ranked or compared with other schools and that all management data collected during the programme implementation would be anonymised. Therefore, there was no incentive for treatment schools to engage in unethical and illegal behaviour involving artificially and systematically altering their students' scores.

III.E. Econometric Specifications

The general management of a school after treatment is the simple average of the scores reached in each of the 23 practices in our December 2023 survey. The outcome variable M_s can represent one specific management practice, the average of a group of management practices or all 23 management practices of the school s . Let Z_s be a dummy that takes a value of 1 if the school s was assigned to the treatment group and 0 if the school s was assigned to the control group. Let n_s be the number of students in each sample school s . As the number of students

per school varies from 156 to 1346, it is important to weigh this school-level analysis with the number of students per school. Therefore, the average treatment effect (ATE), λ_{ATE} , can be identified using the following weighted regression:

$$M_s\sqrt{n_s} = \iota\sqrt{n_s} + \lambda_{ATE}Z_s\sqrt{n_s} + \varepsilon_s\sqrt{n_s}. \quad (1)$$

Equation (1) identifies the ATE of the programme on school management. Standard errors are clustered robust standard errors (Chaisemartin and Ramirez-Cuellar 2024). The attrition rate is null because all the schools participated in the December 2023 survey.

We also aim to estimate the ATE of the SMEP on students' educational outcomes. The analysis is conducted separately by subject. The treatment lasted two years, from January 2022 to December 2023. For each student i , let Y_{isp} represent the outcome of interest—the individual reading or mathematics score from the December 2023 Rio Assessments, standardised by subject and grade with the control group as the reference. Let Z_{isp} be an indicator variable that takes a value of 1 when the student was enrolled (during November/December 2021) in a treatment school for the 2022 academic year and 0 when the student was enrolled (during November/December 2021) in a control school for the 2022 school year. The subscript s represents the school in which the student i was enrolled for the 2022 school year, and p is the pair of schools used for the random assignment that included the school s . Therefore, the ATE, β_{ATE} , can be identified for each subject using the following causal equation:

$$Y_{isp} = \alpha + \beta_{ATE}Z_{isp} + \epsilon_{isp}. \quad (2)$$

The unadjusted regression shown in Equation (2) identifies the programme's ATE on students' educational outcomes for each subject. We use robust SE clustered by school pairs defined by the random assignment, and we do not include pair-fixed effects in the regression (Chaisemartin and Ramirez-Cuellar 2024).

The attrition rates for student scores in our sample are only 1.1% for reading and mathematics scores from the December 2023 Rio Assessments. This is due to the approach of replacing December 2023's missing outcome data with students' standardised scores from previous tests. It is a conservative approach because we used educational outcomes from students exposed to the treatment for less than two years. Moreover, the attrition rate differences between treat-

ment and control were 0.16% for reading and 0.18% for mathematics; both differences were nonsignificant. More details on the approach used to deal with missing outcome data can be found in Section III.C.

The previous equations focused on the average causal effect of the treatment on students' learning and the schools' management. We also want to clarify the impact of the intervention on students' educational achievements through the intensity of the treatment delivered. The variable of interest here is the management level of the schools measured by the December 2023 survey. As the average management variable is continuous, the average causal response theorem (ACR) and its continuous corollary (Angrist and Imbens 1995) provide the path to identify and estimate the causal effect of the treatment's intensity, the management changes driven by the random assignment (the intervention) in each school, on students' learning.

Based on the ACR theorem and its continuous corollary, the IV estimation, using a variable treatment intensity, produces a weighted average derivative along the length of a possibly nonlinear causal response function (Angrist and Imbens 1995). This response function is the relation between student learning and school management. Additionally, comparing the CDF of the endogenous variable (treatment intensity) while turning the instrument on and off helps identify where the action comes from with the instrument. In this context, the IV recovers the average derivative over the range of the school management levels where the instrument (random assignment) shifts the CDF of the endogenous variable (management) sharply.

We assume that our IV is independent of any factor (the independence assumption). We also assumed that improving school management is the only channel for changing students' learning from the instrument (the exclusion restriction). The programme was delivered to treatment school managers who were uninvolved in direct teaching. Consequently, the treatment assignment could only influence students' learning by improving school management, driven directly by the school managers and indirectly by the programme. Furthermore, due to the strict confidentiality agreement, it is unlikely that any action due to the random assignment was directed to the experimental schools apart from the programme directed to treatment schools.

The analysis is conducted separately for each subject: reading and mathematics. Let Y_{isp} and Z_{isp} be the same variables defined when discussing the ATE. Let M_{isp} be the general management level of school s from the school pair p in which student i registered for the 2022

school year (enrolment in November and December 2021). M_{isp} represents a school’s average score for the 23 management practices in our 2023 survey. Let β_{IV} be the causal parameter of interest: the impact of the treatment intensity (school management changes driven by the random assignment) on students’ educational outcomes. β_{IV} can also be interpreted as the causal effect of a 1-point change in the schools’ general management (a 1–5 scale) due to the instrument on students’ educational outcomes. The parameter, β_{IV} , can be identified from the following two-stage least square (2SLS) procedure using the random assignment as the IV:

$$Y_{isp} = \alpha + \beta_{IV} \hat{M}_{isp} + \epsilon_{isp}, \quad (3)$$

where the $\hat{M}_{isp}$ is the fitted M_{isp} from the following first-stage equation

$$M_{isp} = \omega + vZ_{isp} + \xi_{isp}.$$

Standard errors are robust and clustered, with clusters being the school pairs formed for the random assignment and the noninclusion of school pairs’ fixed effects in the regressions (Chaisemartin and Ramirez-Cuellar 2024).

We also discuss the program’s impact on management practices and students’ educational achievements as well as the causal effect of the changes in management practices on students’ learning with full regression adjustment models using baseline covariates (adjusted models), such as student scores from tests and the schools’ management levels as measured before the experiment began. The details are provided in Appendix H.

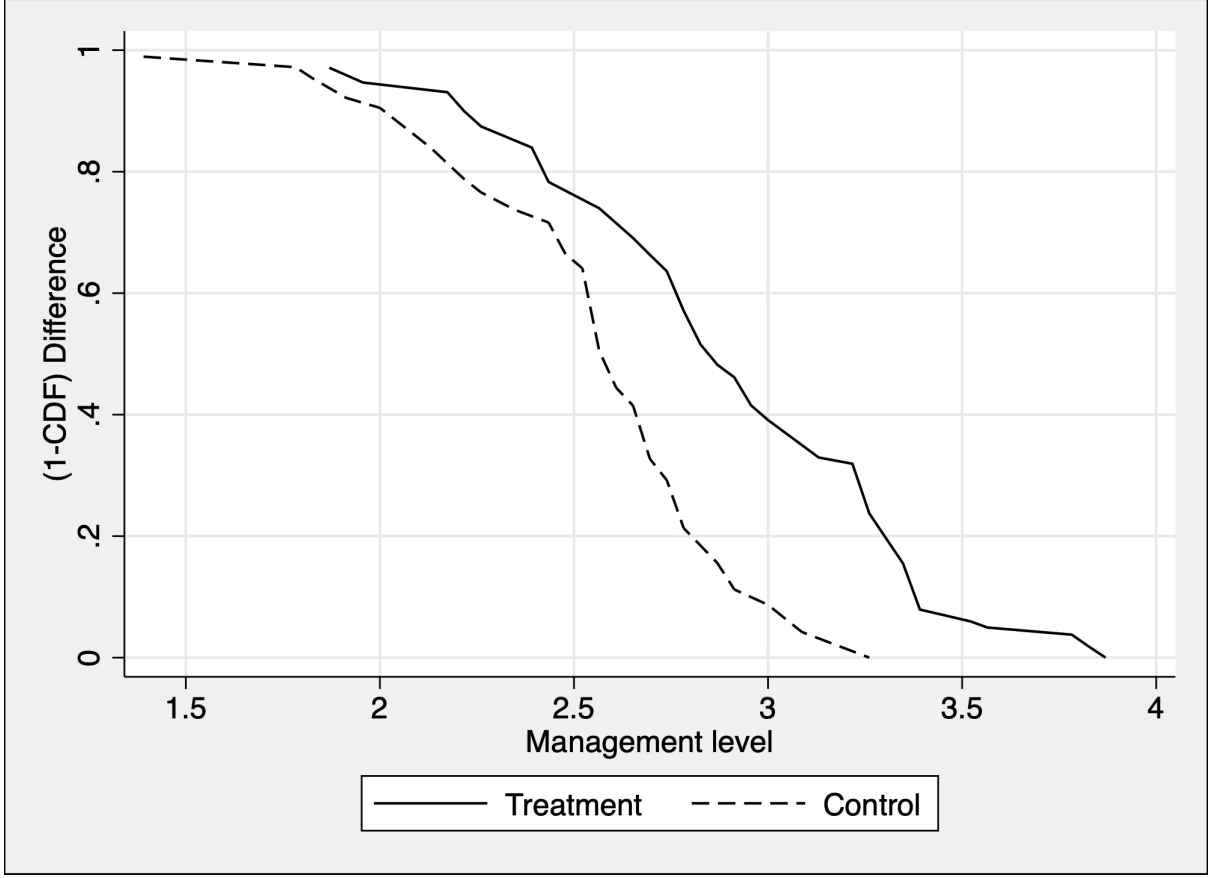
IV. RESULTS

IV.A. *The Impact on Management*

Figure VI demonstrates the impact of our treatment on the schools’ average management level. The figure illustrates the difference between schools in the treatment and control groups in terms of the probability that the schools’ management level, with respect to the 23 best management practices, is greater than or equal to a specific management level indicated on the x-axis. Differences are observed across the x-axis. For example, while the probability that a control school had a management level greater than or equal to 3 is approximately 10%, the

same probability for a treatment school is approximately 40%.

Figure VI: Impact of the Treatment on the Schools' Average Management Level



Notes: This figure illustrates the difference between schools in the treatment and control groups concerning the probability of a school's management level being greater than or equal to a specific management level on the x-axis with respect to the 23 best management practices (i.e., 1 minus the CDF). Although the management level scale ranges from 1 to 5, the x-axis only displays the range where observations are present to improve clarity. The solid and dashed lines represent schools in the treatment and control groups, respectively.

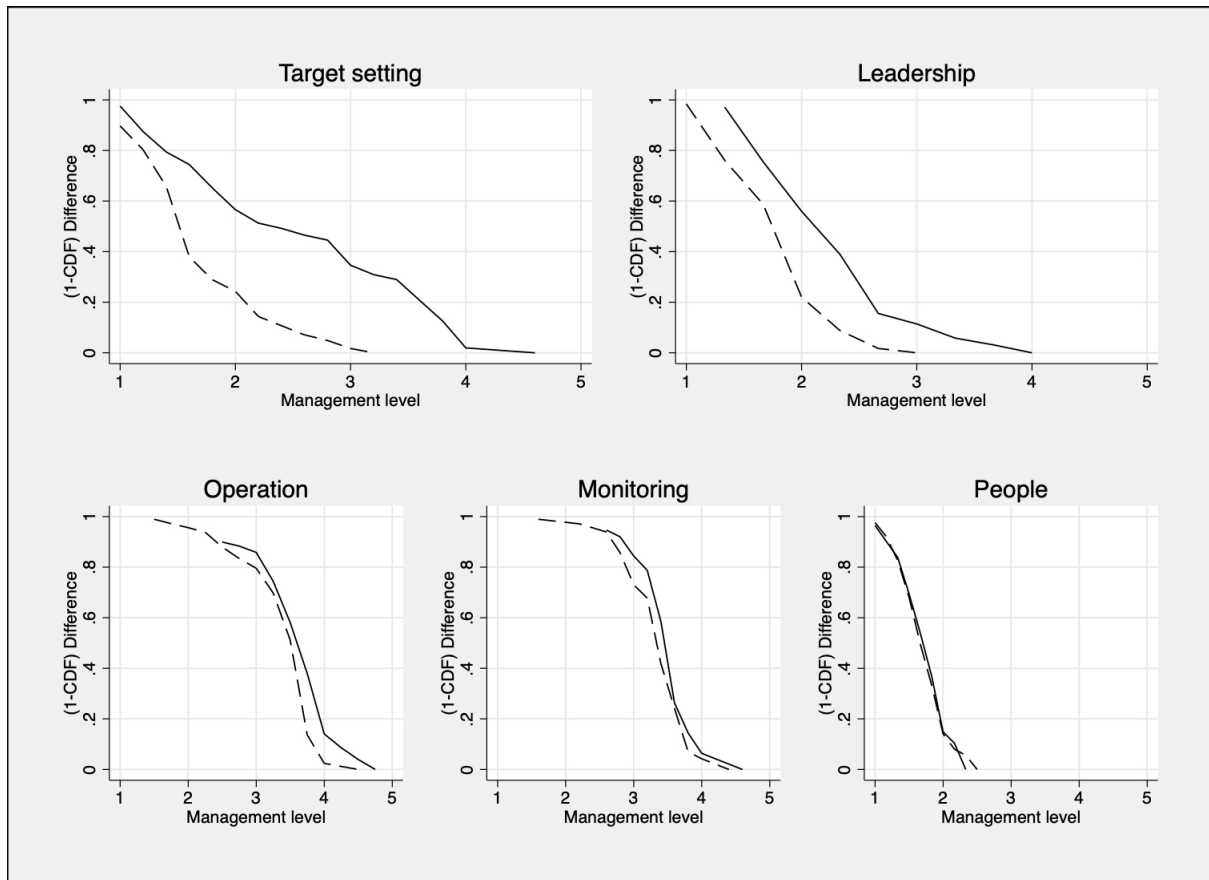
The programme successfully changed the overall management levels of the treatment schools in comparison with the control schools. Given that the 23 best management practices can be grouped into five groups—target setting, leadership, operations, monitoring, and people management—as discussed in Section II.B, it is worthwhile to analyse how the treatment impacted each group.

Figure VII shows the treatment's impact across the five groups. The figure indicates that most of the treatment's effect on the average management level was concentrated in target setting and leadership. For example, while the probability that a school in the control group had a target-setting management practice level of 3 or higher is approximately 0%, among schools in the treatment group, it is approximately 35%. Meanwhile, while the probability that

a school in the control group had an average leadership practices score of 2 or higher is 20%, it is approximately 60% for schools in the treatment group.

In comparison, the differences between the treatment and control schools are smaller in the operation and monitoring groups of practices.²⁵ For example, while the probability that a school in the control group had an operation average score of 3 or higher is approximately 70%, it is approximately 85% for schools in the treatment group. Furthermore, according to Figure VII, no differences were observed between the control and treatment groups with respect to people management.

Figure VII: Impact on the Groups of Management Practices



Notes: The figure displays the difference between schools in the treatment and control groups regarding the probability that a specific management principle group of a school is greater or equal to a specific management level on the x-axis (i.e., 1 minus the CDF). The management level scale on the x-axis ranges from 1 to 5. Solid and dashed lines represent treatment and control schools, respectively.

Table III presents the ATE for general school management practices and the five management practice groups under different specifications. Each dependent variable was standardised to have a mean of 0 and an SD of 1 for the control group. No additional controls were included

²⁵It is relevant to remember that treatment schools' CDF was slightly behind the control schools' CDF regarding operation and monitoring group of practices in 2021, according to Figure IV.

in the weighted regressions reported in column (1). The coefficients reported in column (2) include the regressions in the following centered (demeaned) baseline covariate and its interaction with the treatment variable: the schools' management scores from the 2021 survey. The weights used are the number of students in grades 1–8 in each school, as discussed in Section III.E.

Table III: The Impact of the Treatment on Management

	No covariate	With covariate
	(1)	(2)
	ATE	ATE
	OLS	RA
General Management (SD)	0.916*** (0.268)	0.928*** (0.260)
Management by Group of Practices		
Target setting (SD)	1.530*** (0.364)	1.528*** (0.368)
Leadership (SD)	1.069*** (0.278)	1.079*** (0.272)
Operation (SD)	0.370 (0.223)	0.400* (0.217)
Monitoring (SD)	0.324 (0.232)	0.332 (0.232)
People management (SD)	0.054 (0.227)	0.000 (0.000)

Note: This table shows the causal impact of the treatment on average management level and across different groups of management practices: target setting, leadership, operation, monitoring, and people management. The management level of each of the 80 experimental schools, in relation to the 23 best management practices, was measured by a survey conducted in December 2023, following the WMS methodology. Each practice was scored using a grid ranging from one (worst management) to five (best management). The average general management score is calculated by averaging the school's scores across the 23 best practices, whereas its average score in a specific group is calculated by averaging the scores for practices in that group. The scores were standardised using the control group as a reference. Column (1) reports average treatment effects (ATE) from regressions of the average management scores on a treatment dummy without control variables. Column (2) shows ATE estimates from regressions of the average management scores on a treatment dummy, the centered average score from 2021, and the interaction between the centered baseline covariate and the treatment dummy. Robust standard errors are reported in parentheses, with significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

Table III shows that the average impact of the treatment on school management practices was 0.916 SD for the unadjusted model, and this impact is significant at the 99% confidence

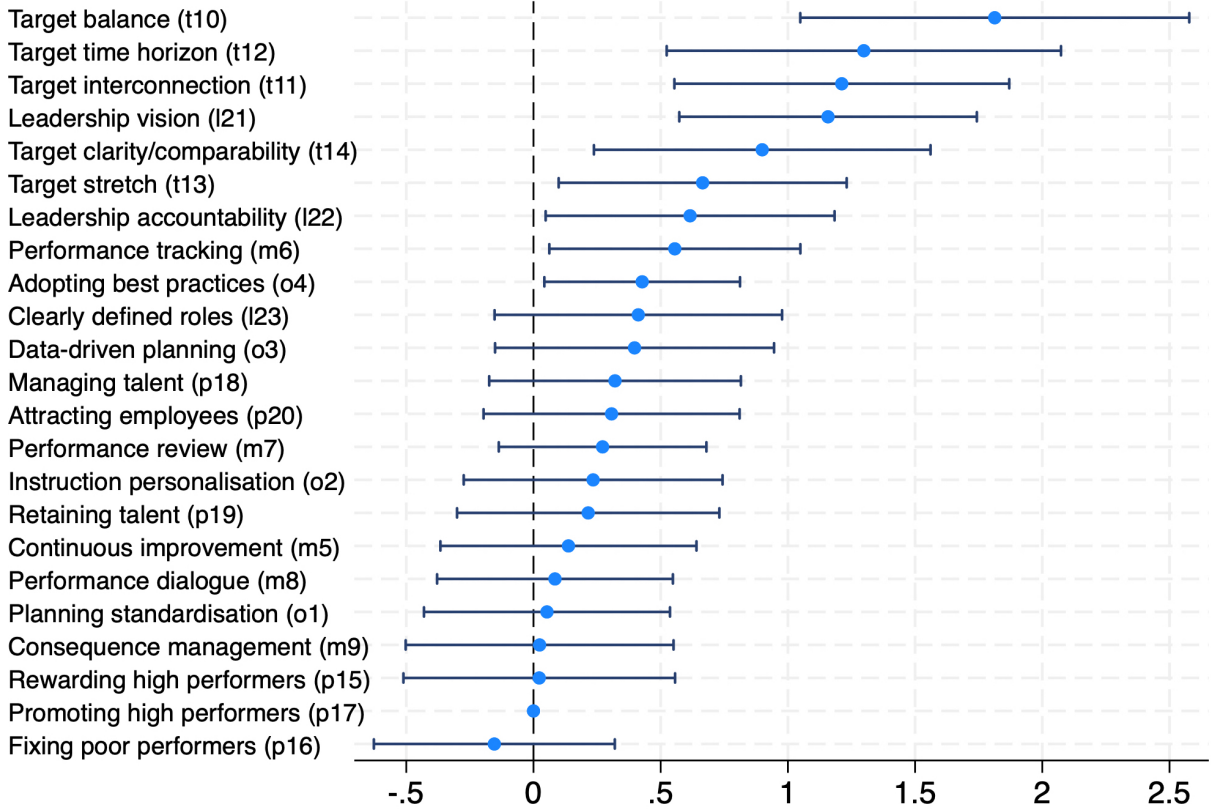
level. The management programme had an effect of 1.530 SD on the target-setting group and 1.069 SD on the leadership group when considering the unadjusted models. Both estimates were statistically significant at the 1% level. The estimated treatment effects on the operation and monitoring groups of practices were 0.370 and 0.324 SD, respectively. However, none of these estimates were statistically significant. Finally, the treatment did not significantly impact the people management group. Quantitatively, the treatment impacts reported in column (2), where we control for the initial level of management practices, are not significantly different from those in column (1).

Thus far, we have reported the impact of our management programme by either averaging all 23 best practices or by grouping them into five different categories and averaging the practices within each category. We also report how the treatment affected each of the 23 management practices. See Figure VIII), which displays the average point estimates and their 95% confidence intervals. Each practice used as an outcome variable in the regressions was standardised to have a mean of 0 and an SD of 1, based on the control group. Thus, the x-axis scale is in SD.

Figure VIII shows that the treatment positively impacted school management across almost all the 23 practices, indicating that the programme successfully achieved its goal. Furthermore, the ATE varied across these 23 practices. The average impact of the treatment was statistically significant at the 5% level and larger than 1 SD in four practices: target balance (T10), target time horizon (T12), target interconnection (T11), and leadership vision (L21). The ATE estimates were statistically significant at the 5% level and between 0.4 and 0.9 SD on five different practices: target clarity/comparability (T14), target stretch (T13), leadership accountability (L22), performance tracking (M6) and adopting best practices (O4).

Although not statistically significant, the treatment's impact was positive on the following practices: clearly defined roles (L23), data-driven planning (O3), managing talent (P18), attracting employees (P20), performance review (M7), instruction personalisation (O2), retaining talent (P19), continuous improvement (M5), performance dialogue (M8) and planning standardisation (O1). Four practices did not change from the treatment: consequence management (M9), rewarding high performers (P15), promoting high performers (P17), and fixing poor performers (P16). Therefore, the results presented in this figure, with a disaggregated analysis of the 23 management practices, confirm the findings from the grouped analysis reported in Figure VII.

Figure VIII: Impact on Each Management Practice



Notes: The figure shows the average program's impact (ATE) on each of the 23 best management practices. The figure displays the ATE point estimates and their 95% confidence intervals. Each point estimate and confidence interval comes from an unadjusted regression (school level) of one of the 23 management practices on a treatment dummy that takes the value of 1 when the school was randomly assigned to the treatment and 0 otherwise. Robust SE were used. Each of the 23 management practices used as outcome variables in the regressions was standardized to have a mean of 0 and an SD of 1 regarding the control group. Thus, the x-axis scale is in SD.

IV.B. The Impact on Learning

Having documented that the SMEP had a causal impact on school management, we now investigate whether better management practices improved learning. It is important to emphasise that there were no changes in staff, payment rewards, or teaching time; the programme primarily introduced management tools to existing school principals and coordinators.

Table IV reports the impact of our treatment on learning under different specifications. All outcome variables were generated in December 2023, two years after the beginning of the programme implementation. The sample includes all 31,760 students enrolled in grades 1–8 during the 2022 school year in one of the 80 randomly selected schools. These students registered at a sample school in Rio de Janeiro during the regular enrolment period, from November to December 2021, before the experiment began. The students' reading and mathematics scores

(IRT) from the December 2023 Rio Assessments represent their learning. The scores were standardised to have a mean of 0 and an SD of 1 for each grade and subject in relation to the control group. “Mgmt ([1,5])” represents the general management of each school, denoting the simple average of the 23 scores reached by a school for the 23 management practices in our December 2023 survey. Each school’s “Mgmt ([1,5])” ranges from 1 to 5.

Table IV: The Treatment’s Causal Effects

	No covariates		With covariates	
	(1)	(2)	(3)	(4)
	ATE	Mgmt ([1,5])	ATE	Mgmt ([1,5])
	OLS	IV	RA	IV/RA
Reading (SD)	0.226*** (0.059)	0.680*** (0.245)	0.187*** (0.045)	0.564*** (0.182)
Maths (SD)	0.237*** (0.059)	0.714*** (0.265)	0.208*** (0.049)	0.627*** (0.212)

Note: The table shows the impact of the treatment, SMEP, under different specifications. All outcome variables are from December 2023, two years after the beginning of the treatment. The sample includes all students (31,760) enrolled in grades 1 to 8 at the beginning of 2022 in one of the 80 randomly selected schools from the Rio de Janeiro population of schools. Before the experiment began, these students registered to a sample school in the regular enrolment period in November and December 2021. The students’ reading and mathematics scores (TRI) from the December 2023 Rio Assessments represent their learning. The scores were standardised to have a mean of zero and a standard deviation of one for each grade and subject in relation to the control group. “Mgmt ([1,5])” represents the schools’ average management score that goes from 1, worst management, to 5, the best management. “Mgmt ([1,5])” is derived from a survey conducted in December 2023 that adopted the World Management Survey (WMS) methodology. From 31,760 sample students, there are 31,412 and 31,405 valid students’ scores for reading and mathematics, respectively. Columns (1) and (3) report average treatment effects (ATE) estimates from regressions without and with baseline covariates, respectively. Columns (2) and (4) show IV estimates generated from regressions without and with baseline covariates, respectively. IV estimates can be interpreted as the causal effects of a 1-point change (on a 1–5 scale) in school management, driven by the random assignment/treatment, on students’ educational outcomes. Columns (3) and (4) report ATE and IV estimates from regressions with the following centred baseline covariates (and their interaction with treatment): students’ reading and mathematics scores from previous Rio Assessments that were applied in April and June 2021, and schools’ average management score from our 2021 management survey. Clustered robust standard errors are reported in parentheses. The clusters for these standard errors are the school pairs used for the random assignment procedure. Significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

No covariates were included in the regressions that generated the estimates reported in columns (1) and (2) of Table IV. The coefficients shown in columns (3) and (4) are from regressions that include the following centered (demeaned) baseline covariates and their interactions with the treatment variable: students’ reading and mathematics scores standardised by grade, using the control group as the reference, from two Rio Assessments conducted in April and June

2021. From a total of 31,760 students, 31,412 and 31,405 valid students' scores for reading and mathematics, respectively, were obtained. All 80 sample schools have valid management scores.

Column (1) of Table IV reports the ATE on students' learning outcomes, represented by the treatment effects on reading and mathematics. The ATE estimates are 0.226 and 0.238 SD for reading and mathematics, respectively. The coefficients are statistically significant at the 99% confidence level. Column (3) shows slightly lower ATE estimates from regressions with centred baseline covariates and their interactions with the treatment.

The estimated causal effects of improvements in school management practices, generated by the random assignment, on student learning are reported in columns (2) and (4) of Table IV. In column (2), without covariates, the IV estimates of the impact of a one-point change (on a 1–5 scale) in school management, driven by the treatment, are 0.680 and 0.714 SD for reading and mathematics, respectively, which are both significant at the 1% level. Column (4) reports slightly lower IV estimates from regressions with centered baseline covariates and their interaction with the treatment; however, the effects are also statistically significant at the 99% confidence level.

Section III.C reports that a typical student in a control school gained, on average, 13.14 and 14.73 points in reading and mathematics, respectively, per grade (and year) based on the December 2023 Rio Assessments. Additionally, the average SD by grade was 45.30 for reading and 42.50 for mathematics.²⁶ Therefore, each year of schooling in a typical school in the control group provided, on average, a 0.290 SD ($= \frac{13.14}{45.30}$) of learning in reading and a 0.343 SD ($= \frac{14.73}{42.50}$) of learning in mathematics.

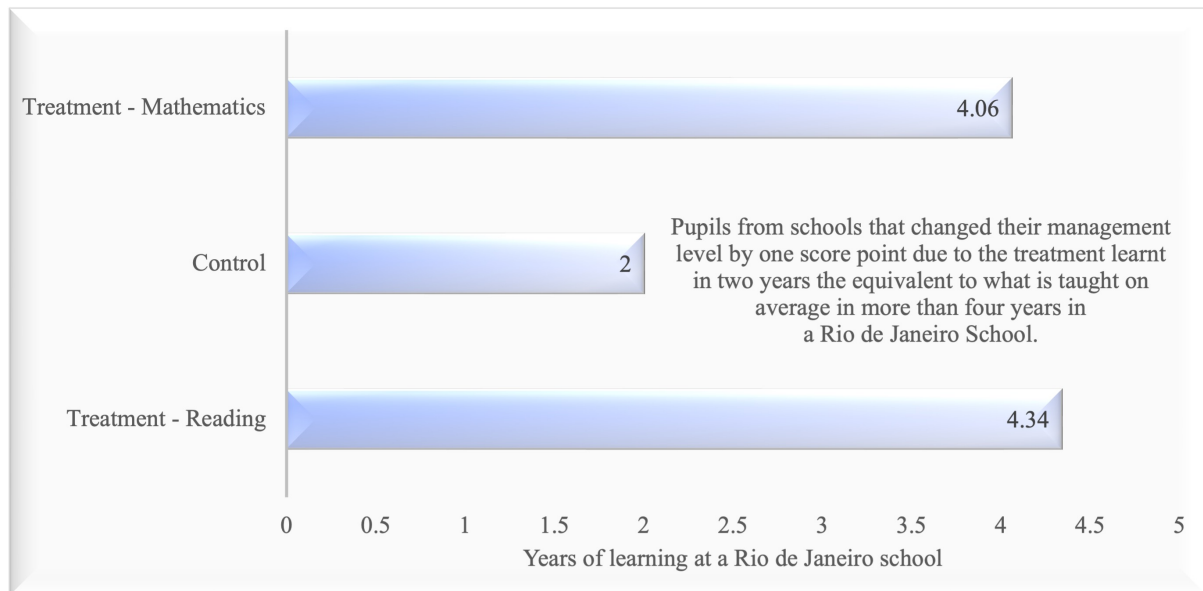
Consequently, the ATE estimates in terms of years of learning in a Rio de Janeiro school are $\frac{0.226SD}{0.29SD} = 0.78$ years for reading and $\frac{0.237SD}{0.347SD} = 0.68$ years for mathematics after 2 years of the implementation of the programme. Therefore, students in schools in the treatment group are, on average, approximately three-quarters of a school year ahead of their counterparts in the control group in terms of learning acquired in reading and mathematics as of December 2023. Treatment schools are, on average, 39% and 34% more productive in providing learning in reading and mathematics than control schools, respectively.

Figure IX shows the IV effects of the impact of high-implementation schools—schools that changed their management level by around one point (on a 1–5 scale) due to the treatment—on

²⁶We calculate the SD within each grade by subject and then average them by grade for each subject.

students’ educational performance. The “Control” bar represents the average learning acquired by students in two years at control schools. The IV estimates are $\frac{0.680SD}{0.290SD} = 2.34$ years of learning for reading and $\frac{0.714SD}{0.347SD} = 2.06$ years of learning for mathematics.

Figure IX: Management Impact (IV) in Terms of Years of Learning



Notes: The figure shows the IV estimates of the impact of high-implementation schools—schools that changed their management level by around one point (on a 1–5 scale) due to the treatment—on students’ educational performance. The “Control” bar represents the learning acquired on average by control students in two years. Students in high-implementation schools are more than two academic years of learning (reading and mathematics) ahead of students from schools that did not receive the treatment. High-implementation schools taught students reading and mathematics skills equivalent to what is learned by spending more than four years in a control school. Schools in the treatment group were 117% more productive in providing learning in reading and 103% in mathematics than schools in the control group.

Students in high-implementation schools are approximately two academic years of learning (in reading and mathematics) ahead of those from schools that did not receive the treatment. Treatment schools were 117% more productive in providing learning in reading and 103% in mathematics than nontreatment schools.²⁷

IV.C. The Causal Effects on Pupils’ Absenteeism

The SMEP also had an impact on student absenteeism. This effect is based on regressions of the student absence rate on a treatment dummy, with robust SE clustered by school pair.

²⁷It is possible to identify these schools if we consider the impact of treatment on management within each pair of schools used for the pair-matching. We run a regression of the overall school management on a treatment dummy for each pair. For practical purposes, we define a high-implementation school as a treated school that changed its management level by 0.5 points or higher than its paired control school. On average, 19 schools changed their overall management by around one point due to the treatment compared with the control schools within their pairs.

The treatment led to a 19% reduction, statistically significant at the 99% confidence level, in student absenteeism in 2023 compared with students in control schools.²⁸ In 2022, students in treatment schools were 18% less absent than those in control schools. By contrast, in 2021, students in treatment and control schools had nearly the same absence rate, with a difference of 0.2%, which was not statistically significant. Notably, there were no significant differences between the control and treatment groups in terms of student dropout rates, with approximately 0.37% of the students in both groups dropping out in 2023.

IV.D. Heterogeneous Treatment Effects

In this section, we investigate the heterogeneous impact of the treatment on student learning across different groups: gender; race; whether students' households are beneficiaries of Bolsa Familia, the Brazilian cash transfer programme; school grade (1—8) in 2022, school segment, school size, school management in 2021, and the quartiles of student educational outcomes in the June 2021 Rio Assessments. Each of these groups was analysed separately with a regression that included the centred covariate representing the subgroup and its interaction with the treatment dummy. Robust SE were clustered by school pairs. The outcome variable was generated by averaging the reading and mathematics standardised scores from the December 2023 Rio Assessments. The separate regressions for reading and mathematics produced similar results. Scores were standardised to have a mean of 0 and an SD of 1 for each subject, grade, and test application date, using the control group as the reference. The students' performance quartiles were defined using the average standardised reading and mathematics scores from the June 2021 Rio Assessments.

The differences in the treatment's impact were minor and not significant for all groups, with one exception. Based on our 2021 survey, students from schools with better management practices in 2021 seemed to benefit more from the programme. The treatment's impact on student learning was 0.174 SD larger for students from schools with a 1 SD higher management score in 2021.

²⁸The absence rate was 9.16% and 7.5% for control and treatment schools, respectively.

IV.E. Robustness Checks

Table A7 in Appendix J shows the causal effects when the outcome variable of Rio Assessments scores based on IRT is replaced by Rio Assessments scores based on the CTT. The CTT scores are simply the rate or proportion of correct answers. The estimates reported in Table A7 in Appendix J are very similar to those based on IRT scores reported in Table IV. All results remain statistically significant at the 1% level.

Our primary analysis of the treatment’s impact on learning used student-level data, with SE clustered by school pairs defined by the random assignment (Chaisemartin and Ramirez-Cuellar 2024). We then assessed the treatment’s impact on learning at the school level. Each school’s educational outcome is represented by the average standardised scores of its students in reading and mathematics. We used the number of students in the sample as weights in our model, as the schools vary in size, and estimated the impact of the programme (ATE) on students’ educational outcomes (Kahan et al. 2023). Students’ reading and mathematics scores in the December 2023 Rio Assessments were standardised by grade and subject with the control group as the reference. We used robust SE. The estimates from the school-level analysis are very close to the ATE and IV estimates conducted at the student level and shown in Table IV. They remain statistically significant at the 1% level.

We also use Anderson (2008) to compute the sharpened false discovery rate (FDR) q-values to address testing issues with multiple hypotheses. The FDR represents the expected proportion of rejections that are type I errors (false rejections). Although this method does not account for correlations between p-values, Anderson (2008) demonstrates that the method works well if the p-values are positively correlated, as in our analysis. For instance, if the treatment improves students’ reading scores, it is likely to improve their mathematics scores; similarly, if the treatment enhances target balance (t10), it is likely also to improve target interconnections (t11). Table A8 in Appendix J shows that all significant results presented in this study remain significant under the computation of sharpened FDR q-values.

Finally, although we cannot directly test the exclusion restriction assumption, we provide evidence supporting its validity using a subgroup less affected by the instrument (random assignment). According to the heterogeneity analysis conducted in section IV.D, the lower the school management level in 2021, the smaller the impact of the instrument in 2023. A regression analysis (first stage) of the 2023 school management on the instrument, as discussed in

section III.E, but limited to the subgroup of schools that had a management level in 2021 less than 2, reveals that the instrument had no impact on the management of these schools in 2023, indicating a null first stage.

Moreover, a regression analysis of student educational outcomes on the instrument, as discussed in section III.E, again limited to the subgroup of schools with a management level in 2021 less than 2, shows a null result. Consequently, this no-first-stage sample provides no evidence that the exclusion restriction was violated.

V. DISCUSSION

V.A. Understanding the conflicting literature

The seminal RCT conducted by Bloom et al. (2012a) revealed the large effects of management practices on the productivity of manufacturing firms in India. Moreover, Bloom et al. (2020) documented that even nine years later, differences in managerial practices persisted between treatment and control groups, confirming the lasting influence of management on firm productivity. A randomised field experiment to improve management performed by Bruhn, Karlan, and Schoar (2018) involving 432 small- and medium-sized enterprises in Mexico found that treatment firms increased productivity, return on assets, and “entrepreneurial spirit” among the owners. The researchers also used Mexican social security data to identify a sustained 50% rise in both employee numbers and total wage expenditures five years post-intervention. Additionally, Gosnell, List, and Metcalfe (2020) showed that specific management practices, such as monitoring and target-setting, significantly increased aeroplane captains’ productivity on targeted fuel-saving, reducing CO2 emissions. Moreover, captains who participated in the treatment reported higher job satisfaction.

Evidence strongly suggests that management can boost firm productivity, yet causal findings on how managerial practices affect student learning are conflicting and unclear. Our large-scale randomised field experiment findings solve the contradictory evidence on the effectiveness of specific management practices to improve student learning by reinforcing the once-unclear positive evidence and clarifying the null effect literature. More about the previous unclear positive evidence can be found in Appendix A. Let us focus on the null effect literature.

First, a large-scale experimental evaluation conducted by Muralidharan and Singh (2020)

showed no effect of randomly providing 1,774 schools with comprehensive assessments, detailed school ratings, or customised school improvement plans. Notably, the programme did not focus on directly changing the schools’ management through support or accountability measures. As a result, the authors reported no oversight or accountability increase in schools due to the treatment. The study also reported no change in school management or student achievement. This experiment is a remarkable example of managers not necessarily applying the relevant information provided to them. Management programmes that focus only on providing relevant information or training to school managers may not succeed in achieving change in organisations’ management practices.

Second, an RCT conducted by Hoyos, Ganimian, and Holland (2019) in La Rioja, Argentina, reported a large effect of providing schools with students’ educational diagnostic feedback reports and some support (treatment one). However, the study shows no impact (compared to treatment one) in providing schools with professional development workshops on school management and school visits with diagnostic feedback and some support (treatment two). The authors suggested that the lack of statistically significant differences between providing more (treatment two) and less (treatment one) support on school management would represent a null effect of more “management” on student learning. However, the difference between treatments one and two was very weak regarding management. The support given to schools from treatment two, which was not delivered to schools from treatment one, only comprised one visit per year, two workshops in the first year, and three workshops in the second year. Based on Bloom et al. (2012a, 2015), we need to change the overall school management by 1 SD to change student learning by 0.24 SD. We surmise that only large changes in the general management of an organisation can affect productivity consistently.

Third, Romero et al. (2022) conducted a large-scale experiment in Mexico by providing schools with two random treatments: direct management training and indirect (cascade-style “train the trainer” model) training. The study shows that direct training improved managerial capacity by 0.13 SD but did not affect student educational achievement. As we outlined before, small management changes may not be enough to improve educational outcomes or to generate an improvement that could be captured by econometric analysis. The small change in management reported by the study can be explained by the fact that only training school managers may fail to achieve change in organisations’ management practices. It is crucial to ensure that

the practices are changing within the organisation.

V.B. Why Haven't Public Schools Implemented the Best Management Practices?

The programme's implementation over two years, along with the two comprehensive surveys, provides valuable insights into the challenges school managers face in Brazil. The findings suggest that a lack of knowledge and incentives are key factors contributing to the persistence of poor management practices in schools, aligning with evidence from management theorists Gibbons and Henderson (2012). School managers are often overwhelmed by day-to-day bureaucratic tasks and challenges. Although civil society, the media, and watchdog organisations, such as courts of accounts, exert strong pressure on school managers to prioritise maintaining and repairing building facilities, ensuring student meals, and providing security, managers have much weaker incentives to adopt management practices aimed at enhancing student learning.

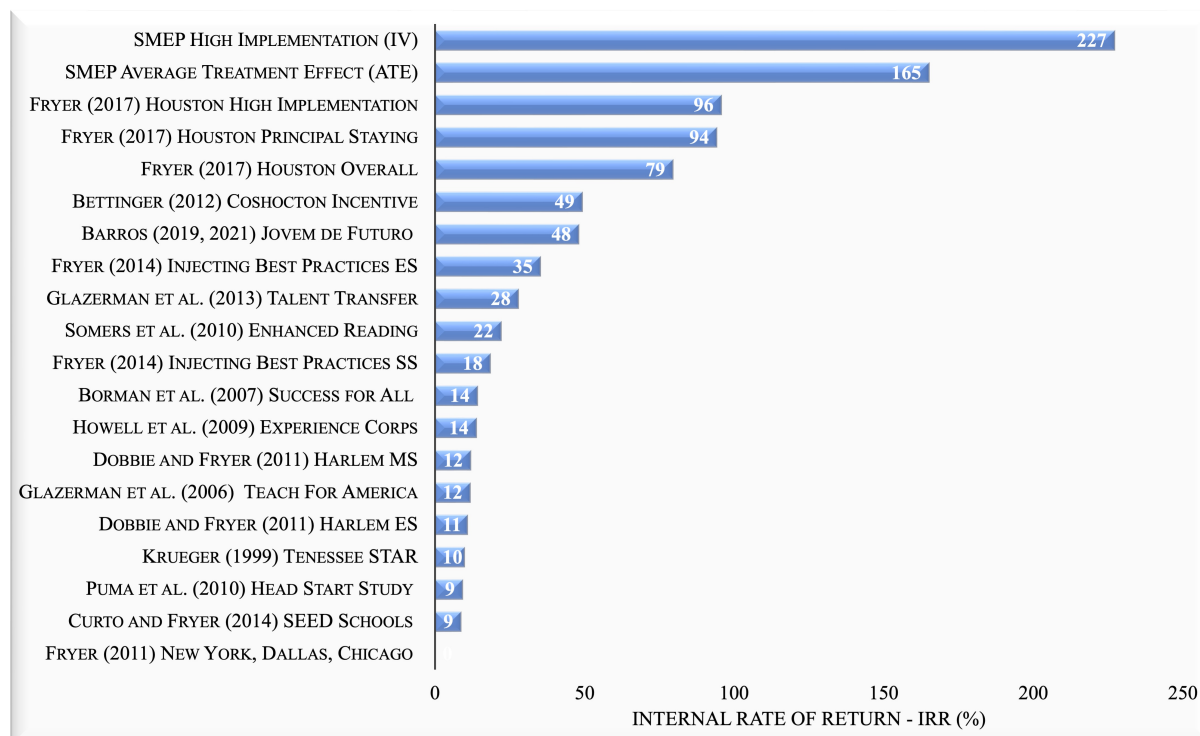
The impact of the SMEP on student learning can be explained as follows. First, the programme provided detailed information to school managers regarding specific management practices customised to meet their needs by the implementation team, helping school managers bridge a knowledge gap about the best practices to enhance student learning. Second, the intervention gave strong support *in loco* to school managers to implement the best managerial practices. Working *in loco* with school managers made it possible for the implementation team to ensure that the practices learnt were being applied to the schools. The process was as simple as possible, aligning with findings from substantial evidence on the positive impact of “make it easy” principles on programme participation (Thaler and Sunstein 2021). The programme succeeded in bringing school managers' attention to ways, through better management, to improve pupils' learning. In conclusion, the intervention filled the information and incentive gaps necessary to overcome the persistence of poor management practices in schools.

V.C. Cost-Benefit Analysis

Figure X shows that the SMEP is highly cost-beneficial compared with other important interventions in the literature. Figure X shows the internal rates of return (IRRs) for our treatment, the “SMEP ATE,” the high implementation of our treatment, “SMEP High Implementation (IV),” the Brazilian programme Jovem de Futuro analysed in Barros et al. (2019) and Barros et al. (2021), and 17 important causal studies discussed in Fryer (2017a,b). The

IRR is the discount rate equating the cost of an intervention with the present value of the expected future earnings inflows generated by the intervention. We calculated the IRRs following Krueger (2003). More details are provided in Appendix K.

Figure X: Internal rates of return (%) for relevant interventions in education



Notes: The figure shows the IRR for our treatment, “SMEP ATE,” the high implementation of our treatment, “SMEP High Implementation (IV),” the Brazilian programme Jovem de Futuro discussed in Barros et al. (2019) and Barros et al. (2021), and 17 important causal studies discussed in Fryer (2017a,b). Apart from our experiment and the Jovem de Futuro’s programme, all other studies’ IRRs come from Fryer (2017a). We calculated the IRR for the SMEP and the Jovem de Futuro Programme following Krueger (2003).

Table A9 in Appendix K helps explain why the SMEP has the highest internal rate of return among the interventions reported in the table. Our programme has one of the lowest per-student costs and the highest impact on student learning (sum of the impact in reading and mathematics). For instance, the SEED Schools intervention had a large impact of 0.42 SDs (adding reading and mathematics estimates); however, it was accompanied by a very high cost of USD 51,904.45 per student per year compared with USD (PPP) 15.22 in our programme.

V.D. Cost-Effectiveness Analysis

We also follow the J-PAL approach, as presented in Dhaliwal et al. (2013), to calculate and compare our treatment’s cost-effectiveness with those of other interventions in low- and

middle-income countries. More details of the J-PAL approach are provided in Appendix L.

Table A10 in Appendix L presents the cost-effectiveness of our programme, the “SMEP ATE, Brazil,” the high implementation of our program, the “SMEP High Implementation (IV), Brazil,” and 30 randomised interventions on education analysed by J-PAL (2020) in low- and middle-income countries. The high-implementation schools of our programme (IV estimates) achieve a cost-effectiveness of 4.80 “additional SDs per \$100.” Based on the ATE estimates, the cost-effectiveness of our treatment reaches 1.71 “additional SDs per \$100.” The SMEP is comparatively one of the most cost-effective programmes using this methodology.

V.E. Low-cost of the intervention

A crucial factor in understanding the significant results achieved by our intervention in the cost-benefit and cost-effectiveness analyses is its low annual cost of USD (PPP GDP) 15.22 per student. The following factors can explain the low cost of our treatment: the intervention was implemented without increasing teaching time, changing existing systems or personnel or providing any financial incentives; and our team only interacted with school managers without any work conducted with teachers or students. These factors limited the programme’s implementation team to a few professionals: six civil servants, one of whom was the manager. More details are provided in the Appendix M.

VI. CONCLUSION

This paper provides experimental evidence of the power of management practices in driving student learning. Given that most interventions are ineffective in improving educational outcomes, with the effective ones tending to be relatively expensive (Fryer 2017a; J-PAL 2020; Kraft 2020; Angrist et al. 2021, 2025), the results of our experiment make a significant contribution to the literature. Our experiment presents a unique intervention, costing only US\$ 15.22 per student per year (adjusted for PPP), that improves school management through the 23 “best” managerial practices while significantly boosting student learning outcomes.

We first show that the SMEP significantly improved overall school management levels. The estimated ATE on school management, measured after two years, was 0.928 SD, a statistically significant result at the 1% level. Additionally, the SMEP had a notable impact on student learning. The ATE for reading and mathematics were 0.226 and 0.237 SD, respectively, with

both results being statistically significant at the 1% level. After two years, students in the treatment group were three-quarters of a school year of learning ahead of their peers in the control group. The schools in the treatment group were 37.5% more productive in delivering learning to their students. The treatment also reduced absenteeism by 19% in the treatment schools compared with control schools.

Finally, our treatment intensity analysis showed that a 1-point improvement in school management had a causal effect of 0.680 SD in reading and 0.714 SD in mathematics. Likewise, a 1-SD change in management had an effect of 0.247 SD for reading and 0.259 SD for mathematics. These estimates are statistically significant at the 1% level. Notably, students in high-implementation schools were over two academic years of learning ahead of their peers in control schools; that is, high-implementation schools were 110% more productive in providing learning to their students than control schools. Therefore, better-managed schools, based on the 23 best management practices, can significantly improve student learning.

Despite differences in income levels, investment in public education, and other factors, Brazil's public schools share similar organisational structures. For example, each school has a principal, a coordinator (or supervisor) and teachers. Furthermore, public schools must follow the same national legal rules. The school structure is standardised, even across countries, according to Dobbie and Fryer (2013), Fryer (2014), and Fryer (2017a). As school management is likely to matter not only for schools in Rio de Janeiro but also for all public schools in Brazil, the findings of this study can have far-reaching implications for Brazil's education policy, especially as the country continues to face significant challenges in delivering high-quality public education. More information regarding the external validity debate can be found in Appendix N.

Although the findings are representative, this experiment is limited to one context and tests a subset of possible management interventions. Similar trials in different contexts would be beneficial for determining how well this approach can be adapted to other countries and organisations, such as hospitals. Furthermore, as the results were collected immediately after the programme's second year, exploring the long-term effects of the intervention on school management, student learning, and other outcomes would be valuable.

Further research could explore incentives to increase the implementation of the best management practice, given this study's finding that the greater the improvement in management

practices, the greater the effects on student learning. Making the programme compulsory seems to be the most intuitive path to increasing participation. Another solution could be to link the higher implementation of the practices with moral incentives, such as an award for the efforts of school management in implementing the practices. Future research could also tackle which of the 23 practices are more critical in driving students' educational outcomes.

Based on the promising results reported by the experiment, Rio de Janeiro has already expanded the SMEP to over 100 schools. The goal is to have the SMEP in all the municipality's 992 grades 1–9 public schools. Other states and municipalities in Brazil have been interested in implementing SMEP in their schools, showing that Brazilian policymakers and politicians recognise the programme's success. This widespread interest in the programme nationwide indicates it might extend far beyond the initially treated schools.

COURT OF ACCOUNTS OF RIO DE JANEIRO (TCMRio), BRAZIL

UNIVERSITY OF CAMBRIDGE, UNITED KINGDOM and GETULIO VARGAS FOUNDATION (FGV), BRAZIL

Appendix

THE IMPACT OF COST-EFFECTIVE MANAGEMENT PRACTICES ON STUDENT LEARNING: EVIDENCE FROM A LARGE-SCALE RANDOMISED FIELD EXPERIMENT

Felipe Galvão Puccioni

Tiago Cavalcanti

A. Previous Unclear Evidence

Causal studies on the positive effects of management on student educational achievement are unclear and ambiguous (Tavares 2015; Fryer 2017a; Barros et al. 2019; Barros et al. 2021; Beg, Fitzpatrick, and Lucas 2023). Fryer (2017a) investigated the causal effect of principals' management training on student educational achievement through a randomised field experiment involving 58 public schools with the worst educational performance reported in Houston, Texas. The experiment started with randomly assigning principals from 28 schools to receive 300 hours of training on lesson planning, data-driven instruction, and teacher observation and coaching. This management training was based on a book written by Bambrick-Santoyo (2018) and the best management practices discussed by the WMS. There were no staff changes or increases in teaching time.

The results reported in the main tables of Fryer (2017a) are positive and statistically significant. However, the estimates from the Appendix tables are close to zero. The main tables are based on analyses that rely on an outcome variable that sums the subjects (i.e., mathematics + reading), while in the Appendix, the analyses were conducted separately by subject. Using the outcome as the sum of the scores of different subjects makes the estimates challenging to interpret and compare. Based on the analysis in the Appendix tables, the ITT estimates from a “pooled” analysis were approximately 0.03 SD for mathematics and 0.046 SD for reading regarding high-stakes test scores. The ITT estimates, also from a “pooled” analysis, for low-stakes test scores were approximately 0.04 SD across subjects (mathematics, science, reading, and social studies). The effects are slightly better for principals predicted to be high implementers, approximately 0.06 SD by subject, and for principals staying for two years in treatment schools, approximately 0.05 SD by subject.

A randomised field experiment conducted by Beg, Fitzpatrick, and Lucas (2023) claimed to show the positive impact of two management interventions on students’ learning. The problem is the same as that encountered with the charter school studies: it is not possible to disentangle the causal effects of “management” from the effect of the teacher training on Differentiated Instruction (DI) since both treatment arms implemented the DI teacher training and the control group did not implement it. The impact on student learning may be due to DI teacher training programme and not due to any “management” changes implemented in treatment schools.

Three large-scale randomised field experiments performed by Barros et al. (2019) and Barros et al. (2021) with 1,400 state-run high schools (upper secondary schools) across nine Brazilian states analysed the impact of the Jovem de Futuro programme on schools’ educational performance. The programme was delivered in three sequential steps. Each programme step lasted three years and reached different federated states. Additionally, the programme had substantial changes at each step (Barros et al. 2019; Barros et al. 2021). The first step of the Jovem de Futuro focused on changing the mindset and culture of management at the school level. The programme was directly delivered by the nongovernmental organisation Unibanco Institute. Financial incentives were provided to treatment schools.²⁹ No management protocols were part of the programme.

From the second step onward, the programme was implemented by the participating education secretariats with support from the Unibanco Institute. In the second step, management protocols and training were provided for school managers. In the third step, the programme was expanded to reach higher levels of coordination through regional and central coordination of the education secretaries. Management training, protocols, and support were provided for schools and regional and central coordination of the education secretaries. From the third step, no financial incentives were provided to treatment schools.

Despite the changes to improve the Jovem de Futuro at each step of three years of implementation, the intervention’s impact was statistically significant and the same across the steps: approximately 0.1 SD for reading and mathematics. On the one hand, the programme’s impact was robust across different populations and settings. On the other hand, since each programme step had the same impact, it is unclear which implemented practices caused the effects: financial incentives (1st and 2nd steps), focus on changing the management mindset and culture without

²⁹<https://www.institutounibanco.org.br/>

management protocols (1st step), management protocols and training to schools (2nd and 3rd steps), management protocols and training reaching regional and central coordination (2nd and 3rd steps), focus on schools only (1st and 2nd steps) or focus on three levels: schools, regional and central coordination (3rd step). Perhaps improving the regional or central coordination management level did not affect student educational outcomes, or perhaps the management protocols implemented did not work.

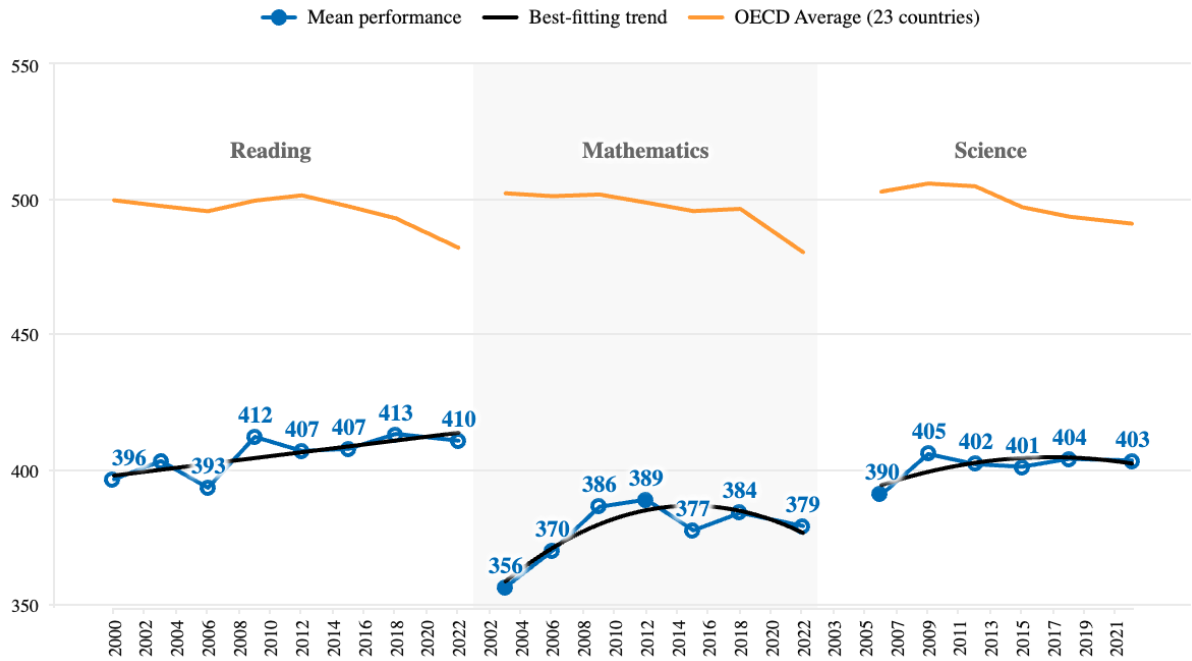
A relevant study based on a fuzzy regression discontinuity design (half fuzzy RDD) conducted by Tavares (2015) showed the positive impact of specific management practices on student learning. The research shows that specific management practices delivered by the Results-based School Management Programme, a management policy developed by the São Paulo Federated State, Brazil, improved the educational performance of 8th-grade students from 0.14 to 0.22 SD in mathematics. However, there was no significant impact on students' reading performance. The management practices delivered by the programme involve management training, strategic planning, goal setting and goal management. Cost information was not available.

B. Efficiency Gap on Education

The OECD's PISA evaluates the aptitude of 15-year-old students in reading, mathematics, and science with respect to real-life problems. Figure A1 illustrates the significant educational gap between Brazil and the OECD member countries for a similar number of years of schooling. The typical 15-year-old Brazilian student scores approximately 72 points lower in reading (482 vs. 410), 88 points lower in science (491 vs. 403), and 103 points lower in mathematics (482 vs. 379), compared with the average 15-year-old student from developed countries in PISA 2022 (OECD 2023).

According to Avvisati and Givord (2021, 2023), Brazilian students' performance in mathematics, reading, and science improves by an average of 12 PISA points over one year of schooling. Dividing the 103-point difference between Brazil's performance in mathematics and that of developed countries in PISA 2022 by Brazil's rate of improvement—12 PISA points per year of schooling—yields the number of years a typical 15-year-old Brazilian student would need to attend a Brazilian school to achieve the performance of a typical 15-year-old student from a developed country.

Figure A1: Brazil's and Developed Countries' Performance in PISA



Notes: The figure compares the PISA performance of Brazil and 23 developed countries from 2000 to 2022. The “OECD Average (23 countries)” yellow lines represent the arithmetic mean performance across the following OECD Member countries: Australia, Belgium, Canada, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Latvia, Mexico, New Zealand, Norway, Poland, Portugal, Sweden, and Switzerland. *Source:* OECD, PISA 2022 Database, Tables I.B1.5.4, I.B1.5.5 and I.B1.5.6. The database can be accessed [here](#).

A typical 15-year-old Brazilian student would need eight and a half more years in a Brazilian school ($8.58 = \frac{103}{12}$) to reach the mathematics performance of a typical 15-year-old student from a developed country. In other words, on average, 15-year-old Brazilian students are eight and a half years behind students from developed countries in mathematics despite the similar number of years of schooling. Meanwhile, Brazil is approximately six years behind in reading ($6 = \frac{72}{12}$) and seven years in science ($7.3 = \frac{0.88SD}{0.12SD}$).

In PISA 2022, 75% of 15-year-old Brazilian students did not achieve a proficiency level of 2 in mathematics. PISA uses a 1–6 proficiency scale. This means that 3/4 of students in Brazil did not meet the more fundamental level of proficiency expected by the end of their first year of high school. After their first year of high school, most 15-year-old Brazilian students cannot interpret simple mathematical situations, such as comparing distances or converting local prices to dollars. Interestingly, in PISA 2003, Brazil had the same proportion of 15-year-old students, 75%, performing below the basic PISA level of proficiency in mathematics (level 2).

The large gap between the educational performance of Brazil and developed countries is not

an isolated case. Almost all developing countries have many years of learning behind developed countries (Pritchett 2013; World Bank 2018; Angrist et al. 2021). Learning inequality also exists within developed countries. The PISA index of economic, social and cultural status (ESCS) reveals that underprivileged students in developed countries scored, on average, more than 100 points lower than their more privileged counterparts with a similar number of years of schooling (OECD 2023). This indicates that underprivileged students are many years behind their more privileged peers, even though they have attended school for the same number of years.

Figure A2 displays the average reading performance (in points) of students in PISA 2018 by country versus the cumulative expenditure per student enrolled in primary or secondary education between the ages of 6 and 15 (in equivalent USD converted using PPP for GDP).³⁰ Straight lines represent the averages of the OECD Member countries in PISA 2018.

The “Overall *efficiency* tendency line” in Figure A2 is a logarithmic tendency line representing the average relationship between reading performance in PISA 2018 and the cumulative per-student spending on education, considering all countries participating in PISA 2018. We use the word *efficiency* because the overall tendency line can also be understood as the average PISA 2018 countries’ efficiency in using educational resources (overall efficiency), that is, education resources vs. educational performance.

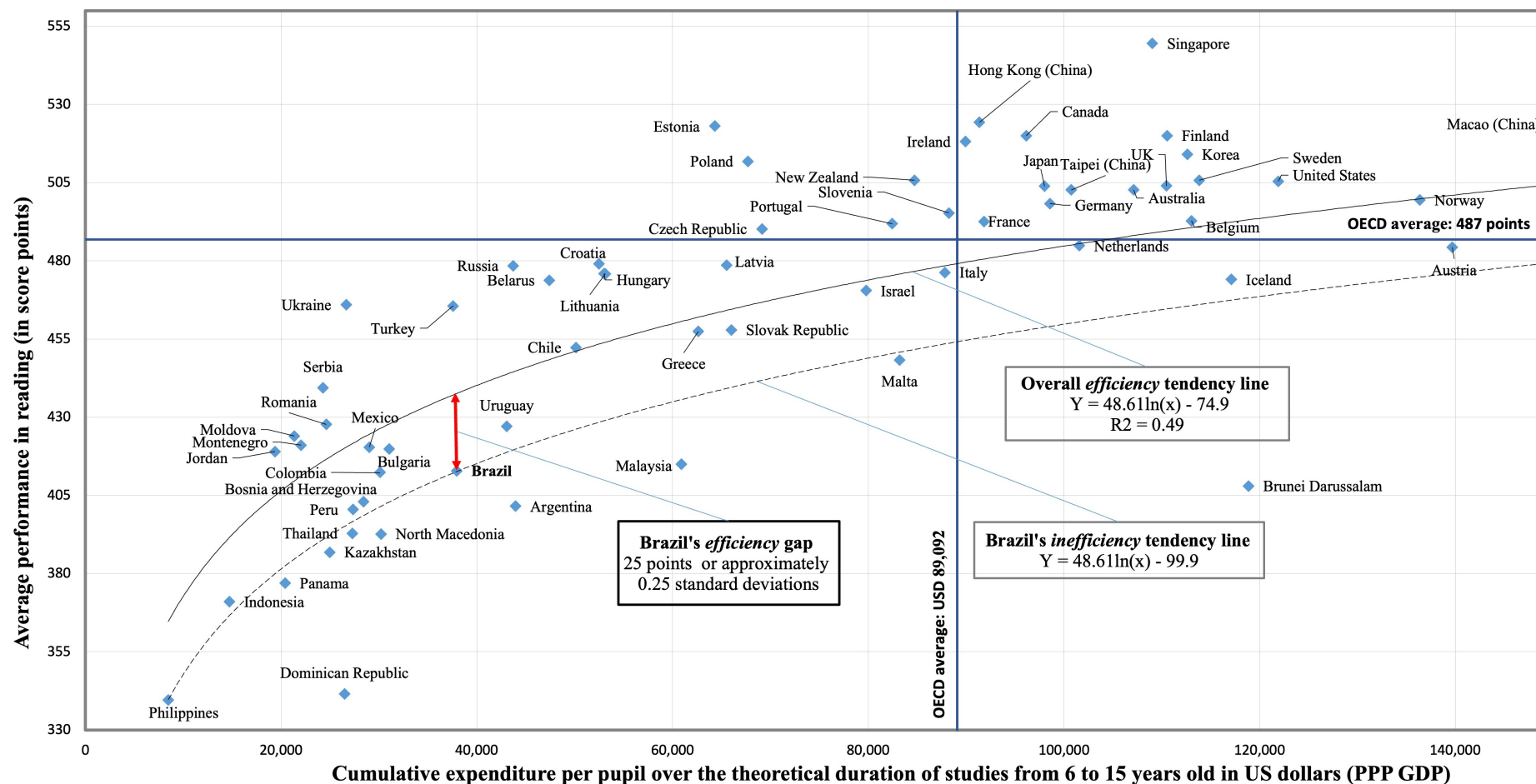
“Brazil’s *inefficiency* tendency line” in Figure A2 is a logarithmic tendency line showing Brazil’s expected reading performance for different cumulative per-student educational spending. The difference between Brazil’s inefficiency tendency line and the overall line is Brazil’s efficiency gap, representing how less efficient Brazil’s spending on education is compared with the average efficiency across all countries participating in PISA 2018.

On average, Brazil reached a reading performance of 413 points for a cumulative expenditure of USD (PPP GDP) 37,954.00, which is 25 points³¹ lower than the expected for its per-student expenditure. For instance, Russia (479 points), Belarus (474 points), Ukraine (466 points), Turkey (466 points), Chile (452 points), Serbia (439), Romania (428), Uruguay (427), Moldova (424), Mexico (420), Bulgaria (420), and Jordan (419) reached higher score points than Brazil in the reading test in PISA 2018 despite similar cumulative per-student spending on education.

³⁰Luxembourg and Qatar are used as data points for the tendency lines (regressions). However, they are not shown in the figure because their large expenditures would make the x-axis much longer, worsening the visualisation.

³¹One hundred PISA score points is approximately 1 SD. Thus, 25 score points are approximately 0.25 SD.

Figure A2: PISA 2018 Reading Performance vs Cumulative Expenditure Per Pupil on Education



Notes: The figure displays the average reading performance (in points) by country in PISA 2018 versus the cumulative expenditure per student enrolled in primary or secondary school between ages 6 and 15 (in equivalent USD converted using PPPs for GDP). *Data source:* PISA 2018 Results (Volume I) - © OECD 2019.

Brazil and Turkey spent approximately USD (PPP GDP) 37,954.00 on education per student from 6 to 15 years old. However, the difference in their proficiency levels (33 points) is quite significant, indicating that Turkey is much more efficient than Brazil in allocating educational resources.

Using the estimates by Avvisati and Givord (2021, 2023), the 53-point difference between Brazil and Turkey indicates that a typical 15-year-old Brazilian student would need to study for more than four years ($\frac{0.53}{0.12}$) at a Brazilian school to reach the typical educational performance level of a 15-year-old Turkish student.

Brazil's *inefficiency* tendency line also shows that Brazil is one of the most ineffective countries in terms of allocating educational resources among the countries in PISA 2018. Brazil's inefficiency in education has also been discussed by Barros et al. (2019) and Barros et al. (2021). Even based on the optimistic assumption that Brazil would follow the same efficiency pattern shown by the overall line, Brazil would not achieve Chile's performance even by doubling its expenditure per student on education. That is, even after spending 50% more than Chile's per-student expenditure, Brazil would still have a lower PISA performance.

Figure A3 shows the performance of OECD countries, Brazil, Turkey and Vietnam, within each quintile of the PISA index of ESCS. ESCS allows students with similar characteristics, such as parent education and family income, to be compared across countries. Quintiles are defined at the international level to include 20% of PISA participants in each quintile. Therefore, the proportion can differ by 20% within each national sample. The size of the markers is proportional to the share of the student population within each quintile of socioeconomic status (as determined by the ESCS). Vertical bars extending beyond the markers represent the 95% confidence interval associated with each estimate. The horizontal, dashed lines represent the 95% confidence interval associated with the mean score of Brazil's largest group of students (as defined by international quintiles).

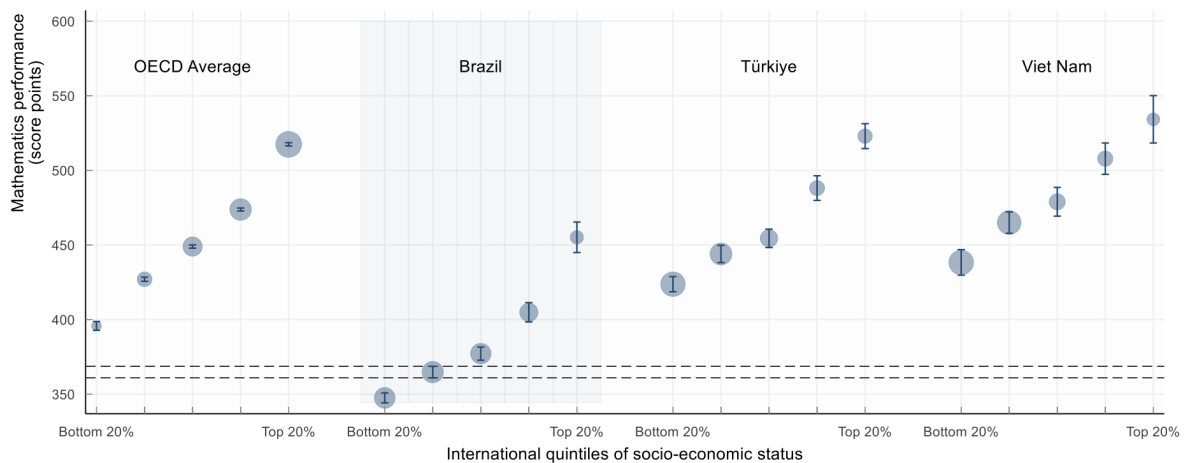
Figure A3 highlights that Brazil's performance in mathematics is much lower than that of the OECD member countries, even when comparing students within the same quintile of socioeconomic status. The figure also compares Brazil with Turkey and Vietnam, which have similar or lower GDP per capita and educational expenditure per student.

The comparison within each quintile of the socioeconomic index shows that Brazil (GDP per capita: USD 8,917.70 in 2022) scored much lower than Turkey, which has a similar GDP

per capita (USD 10,661.20 in 2022).³² Turkey has the same cumulative per student expenditure in education as Brazil, as shown in Figure A2. Even Vietnam, which had a GDP per capita of only USD 4,163.50 in 2022 (half the Brazilian GDP per capita), scored much higher than Brazil in every quintile of the socioeconomic index.

Brazil allocated 6.2% of its GDP to public investment in education in 2018 (Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP) 2021). This is higher than the average of 4.9% among OECD countries in the same year (OECD 2021). The amount invested by Brazil aligns with the educational spending benchmark set by 160 countries in the 2015 Incheon Declaration³³, requiring at least 4%–6% of the GDP to be spent on public education.

Figure A3: Mean performance in mathematics by international quintiles of socioeconomic status



Note: The size of the markers is proportional to the share of the student population within each quintile of socioeconomic status (as determined by the PISA ESCS index). Quintiles are defined at the international level to include 20% of PISA participants in each quintile; within each national sample, the proportion can, therefore, differ by 20%. Vertical bars extending beyond the markers represent the 95% confidence interval associated with each estimate. The horizontal, dashed lines represent the 95% confidence interval associated with the mean score of Brazil's largest group of students (as defined by international quintiles). *Source:* OECD, PISA 2022 Database, Tables I.B1.4.6 and I.B1.4.8.

Brazil's public expenditure on primary, secondary, and postsecondary nontertiary education amounted to 4% of its GDP in 2018, surpassing the OECD average of 3.4% for the same year (Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP) 2021; OECD 2021). Therefore, Brazil's fiscal efforts towards education are comparable with those observed in other OECD countries.

³²Countries' GDP per capita in 2022 can be accessed here.

³³The document can be accessed at <https://unesdoc.unesco.org/ark:/48223/pf0000233137>.

C. Additional Information on Institutional Background

Table A1 compares Rio de Janeiro and other Brazilian municipalities from 2007 to 2019. The per-student public expenditure considers municipality spending on grade 1—9 public schools and nurseries, special education, and young adult education.³⁴ Table A1 is based on data from the Brazilian Education Census³⁵ and 2020 Court of Accounts of Rio de Janeiro (TCMRio) Special Report (p.113).³⁶ All monetary values are updated to December 2019 and are in Brazilian reais (R\$).

Table A1: Education in the Municipality of Rio de Janeiro: 2007 and 2019

Expenditure per student on education			
	2007	2019	Variation
Brazil's cities average	R\$ 4,841.72	R\$ 7,885.57	+63%
Rio de Janeiro City	R\$ 4,805.56	R\$ 9,707.65	+102%

Rio de Janeiro City position - 5th grade Ideb			
	2007	2019	Variation
In RJ Federated State	24th	28th	- 4 positions
In Brazil	1423rd	2529th	- 1106 positions

Rio de Janeiro City position - 9th grade Ideb			
	2007	2019	Variation
In RJ Federated State	7th	28th	- 21 positions
In Brazil	366th	952nd	- 586 positions

Notes: Table based on data from the Brazilian Education Census (can be accessed here) and the 2020 Court of Accounts of Rio de Janeiro (TCMRio) Special Report (p.113) (can be accessed here). All monetary values are updated to December 2019 and expressed in Brazilian reais (R\$). The expenditure per student is calculated by dividing the city's total spending on education by the number of students in nurseries, grade 1–9 schools, and special education plus the number of young adults in special grade 1–9 schools. The Basic Education Development Index (Ideb) brings together, in a single indicator, the results of two equally important concepts for the quality of education: the rate of students' approval and their average performance in reading and mathematics assessments (Saeb). We compare Rio de Janeiro's performance in Ideb with cities within the Rio de Janeiro Federated State and Brazil. RJ Federated State refers to the Rio de Janeiro Federated State.

³⁴The municipality's public expenditure per student is calculated by dividing the municipality's total spending on public education by the number of students in public nurseries, grade 1-9 schools, and special education, plus the number of young adults in state-run grade 1–9 schools.

³⁵It can be accessed at <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-escolar>.

³⁶Access to this report can be found at https://www.tcmrio.tc.br/Noticias/14466/Peca140_RelatorioProjetoParecerPrevio_LAG.pdf.

D. Additional Information about the Science and Management for Education Programme and Its Implementation

Our study complies with all critical research protocols. It received approval from the University of Cambridge’s Ethics Committee and was registered on the AEA RCT registry (AEARCTR-0007669 on May 15, 2021; <https://www.socialscisceregistry.org/trials/7669>). The authorities and institutions responsible for public education in Rio de Janeiro also approved the experiment.

Our intervention aimed to provide the 23 best management practices for schools, as described and discussed by Bloom et al. (2015). These practices can be grouped into five groups: operations, monitoring, target-setting, people management and leadership.

- Operations

- Standardisation of instructional planning processes: the school uses meaningful processes that allow students to learn over time (o1).
- Personalisation of instruction and learning: the school incorporates teaching methods that ensure all students can master the learning objectives (o2).
- Data-driven planning and student transitions: the school uses assessment and easily available data to verify learning outcomes at critical stages (o3).
- Adopting educational best practices: the school incorporates and shares teaching best practices and student strategies across classrooms accordingly (o4).

- Monitoring

- Continuous improvement: the school implements processes towards continuous improvement and encourages lessons to be captured and documented (m5).
- Performance tracking: the school’s performance is regularly tracked with useful metrics (m6).
- Performance review: the school’s performance is reviewed with appropriate metrics (m7).
- Performance dialogue: the school’s performance is discussed with appropriate content and depth and is communicated to teachers (m8).

- Consequence management: mechanisms exist to follow up on performance issues (m9).
- Target setting
 - Target balance: the school covers a sufficiently broad set of targets at the school, department and individual levels (t10).
 - Target interconnection: the school establishes well-aligned targets across all levels (t11).
 - Time horizon of targets: there is a rational approach to planning and setting targets (t12).
 - Target stretch: the school sets targets with the appropriate difficulty level (t13).
 - Clarity and comparability of targets: the school sets understandable targets and openly communicates and compares school, department and individual performance (t14).
- People management
 - Rewarding high performers: the school implements a systematic approach to identifying good and bad performance, rewarding teachers proportionately (p15).
 - Fixing poor performers: the school deals with underperformers promptly (p16).
 - Promoting high performers: the school promotes employees based on job performance (p17).
 - Managing talent: the school nurtures and develops teaching and leadership talent (p18).
 - Retaining talent: the school attempts to retain high-performing employees (p19).
 - Creating a distinctive employee value proposition: the school has a thought-through approach to attract employees (p20).
- Leadership
 - Leadership vision: school leaders have an understanding of the broader set of challenges that the school, system, and key actors face (l21).

- Clearly defined accountability for school leaders: school leaders are accountable for delivery of student outcomes the right mindset to address them (122).
- Clearly defined leadership and teacher roles: the roles, responsibilities, and required attributes of teachers, students, and staff are clearly defined within the school (123).

Figure A4

The 23 Best Management Practices for Schools

 Operations	 Monitoring	 Target Setting	 People Management	 Leadership
1. Standardisation of Instructional Processes School has implemented a clearly defined instructional planning process designed to align instructional strategies and materials with learning expectations and incorporate flexibility to meet student needs; these are followed up on through comprehensive monitoring or oversight. 2. Personalization of Instruction and Learning Emphasis is placed on personalization of instruction based on student needs; school encourages student involvement and participation in classrooms; school provides information to and connects students and parents with sufficient resources to support student learning. 3. Data-Driven Planning and Student Transitions Student transitions are managed in an integrated and proactive manner, supported by formative assessments tightly linked to learning expectations; data is widely available and easy to use. 4. Adopting Educational Best Practices School provides staff with opportunities to collaborate and share best practice techniques and learnings with multiple methods to support their monitored implementation in the classroom.	5. Continuous Improvement Texposing and solving problems (for the school, individual students, teachers, and staff) in a structured way is integral to individual's responsibilities, and resolution involves all appropriate individuals and staff groups; resolution of problems is performed as part of regular management processes. 6. Performance Tracking Performance is continuously tracked and communicated, both formally and informally, to all staff using a range of visual management tools. 7. Performance Review Performance is continually reviewed, based on indicators; all aspects are followed up to ensure continuous improvement; results are communicated to all staff. 8. Performance Dialogue Regular review/ performance conversations focus on problem solving and addressing root causes; purpose, agenda and follow-up steps are clear to all; meetings are an opportunity for constructive feedback and coaching. 9. Consequence Management A failure to achieve agreed targets drives retraining in identified areas of weakness, moving individuals to where their skills are more appropriate.	10. Target Balance Performance metrics and targets are defined for the school and individuals (leaders, teachers, staff) that include both absolute and value-added measures of student outcomes and other metrics linked to key drivers of student outcomes. 11. Target Inter-Connection Goals are aligned and linked at system level and increase in specificity as they cascade, ultimately defining individual expectations for all staff groups. 12. Time Horizon of Targets Long-term goals are translated into specific short-term targets so that short-term targets become a 'staircase' to reach long-term goals. 13. Target Stretch Goals are genuinely demanding for all parts of the organization and developed in consultation with senior staff (e.g. to adjust external benchmarks appropriately). 14. Clarity and Comparability of Targets Performance measures are well defined, strongly communicated and reinforced at all reviews; school performance data includes both quantitative and qualitative measures and are made public.	15. Rewarding High Performers There is an evaluation system which rewards individuals based on performance; the system includes both personal financial and non-financial awards; rewards are awarded as a consequence of well-defined and monitored individual achievements. 16. Removing Poor Performers Repeated poor performance is addressed, beginning with targeted interventions; poor performers are moved out of the school when weaknesses cannot be overcome. 17. Promoting High Performers School actively identifies, develops and promotes its top performing staff members. 18. Managing Talent School proactively controls the number and types of teachers, staff and leadership needed to meet goals; school defines hiring criteria and processes based on understanding of what drives student achievement. 19. Retaining Talent We do whatever it takes to retain our talent. 20. Attracting Talent We provide a unique value proposition to encourage talented people join our school above our competitors.	21. Leadership Vision School leaders define and broadly communicate a shared vision and purpose for the school that focuses on improving student learning and outcomes (often beyond those required by law); vision and purpose is built upon a keen understanding of student and community needs, and defined collaboratively with a wide range of stakeholders; school leader proactively builds environment conducive to learning. 22. Clearly Defined Accountability for School Leaders School leaders are held accountable for quality, equity and cost- effectiveness of student outcomes within the school, with school-level and individual consequences for good and poor performance; leaders are provided sufficient autonomy to impact the areas of accountability. 23. Clearly Defined Leadership and Teacher Roles School defines clear roles, responsibilities and desired competencies of teachers and staff across the school, built upon an understanding of what drives student performance and outcomes; leadership responsibilities are distributed across school.

Source: Since Bloom et al. (2015). defines three scenario scores for each practice: worst (1), middle (3), and best (5); we include each practice in the table with the description of the best school scenario (score 5) for that practice. This table was provided for treatment school managers (principals, deputy principals and pedagogical coordinators).

E. Additional Information on the Management Level of Schools in 2021 and 2023

Table A2 presents the distribution of “subjects” by raters 1 and 2 and the categories for the management surveys conducted in 2021 and 2023. The 1,840 (23 practices x 80 schools) management practices surveyed in each year (2021 and 2023) are the subjects, as the same management practices in different schools serve as different survey “subjects.” In the 2021 survey, 77 subjects were not scored by rater 2 due to problems in the interview recording.

The table shows that raters generally agreed when the scores were low. The percentage of disagreement rose with higher scores.

Table A2: Contingency Tables for the 2021 and 2023 Management Surveys

2021 management survey						
Rater 1	Rater 2					Total
	1	2	3	4	5	
1	409	128	51	11	2	601
2	71	148	113	28	5	365
3	38	106	219	111	12	486
4	5	33	92	99	24	253
5	0	3	22	21	12	58
Total	523	418	497	270	55	1,763

2023 management survey						
Rater 1	Rater 2					Total
	1	2	3	4	5	
1	281	80	6	2	1	370
2	87	195	98	32	2	414
3	13	87	215	153	4	472
4	5	26	148	297	36	512
5	1	1	10	51	9	72
Total	387	389	477	535	52	1,840

Notes: This tables shows the 2021 and 2023 distribution of “subjects” by rater and category. The subjects are the 1,840 (23 dimensions x 80 schools) management practices surveyed in the 80 sample schools, as the same management practice in different schools serves as different “subjects.” In the 2021 survey, 77 subjects were not scored by rater 2 due to problems in the interview recording. Rater refers to the survey team members scoring management practices across the sample schools. Raters scored each of the 23 WMS management practices in each school based on the WMS scoring scale of 1–5.

More systematically, Table A3 shows inter-rater reliability coefficients for the 2021 and 2023 management surveys and a benchmark scale based on Landis and Koch (1977). These coefficients measure the agreement between raters when the ratings are ordinal (Gwet 2021).

The percentage agreement in Table A3 is the only coefficient not corrected by chance agreement; it includes agreements between raters that are and are not due to chance. All other coefficients have paths to exclude rater agreement due to chance from their calculations; they measure the systematic agreement that is not due to chance.

As seen from Table A2, the survey scores are more concentrated in some categories. This phenomenon is called high trait prevalence, and it has a dramatic effect on many inter-rater reliability coefficients, such as Cohen/Conger’s Kappa, Scott/Fleiss’s Pi, and Krippendorff’s Alpha (Feinstein and Cicchetti 1990; Gwet 2021). Consequently, we focus our analysis on the

Brennan-Prediger and Gwet’s AC2 coefficients following Gwet (2021), who finds they are more resistant to high trait prevalence.

Table A3 shows a Brennan-Prediger coefficient of 0.720 and a Gwet’s AC2 of 0.760 for the 2021 survey. The analysis of the 2023 survey presents higher agreement coefficients, such as a Brennan-Prediger coefficient of 0.788 and a Gwet’s AC2 of 0.811. These values are considered by Landis and Koch (1977) to be relatively large and reflecting close to perfect inter-rater agreement. The high agreement coefficients in both surveys indicate that the scores assigned to each of the 23 management practices for each school are independent of the specific survey team member (rater or coder) who evaluated the school. Consequently, the evidence suggests that the management scores given to schools are not the raters’ subjective views but rather actual school characteristics.

Table A3: Agreement Coefficients for the 2021 and 2023 Management Surveys

	2021	2023
Percent Agreement	0.922	0.941
Brennan and Prediger	0.720	0.788
Cohen/Conger’s Kappa	0.608	0.709
Scott/Fleiss’ Pi	0.608	0.709
Gwet’s AC2	0.760	0.811
Krippendorff’s Alpha	0.608	0.709

Benchmark scale	
<0.000	Poor
0.000 - 0.200	Slight
0.200 - 0.400	Fair
0.400 - 0.600	Moderate
0.600 - 0.800	Substantial
0.800 - 1.000	Almost Perfect

Notes: This table shows different inter-rater reliability coefficients for the 2021 and 2023 management surveys. They measure the agreement between raters when the ratings are ordinal (Gwet 2021). The WMS management score is rated from one to five. Two raters scored each of the 23 management practices for each school in both surveys. As there were 80 schools and 23 dimensions, 1,840 “subjects” were scored by two raters in each survey (2021 and 2023). The percent agreement is the only coefficient not corrected by chance agreement. The Brennan–Prediger and Gwet’s AC2 coefficients are presented in bold because they are robust to high trait prevalence paradoxes such as those encountered in the scores from our surveys (Gwet 2021). The benchmark scale was developed by Landis and Koch (1977).

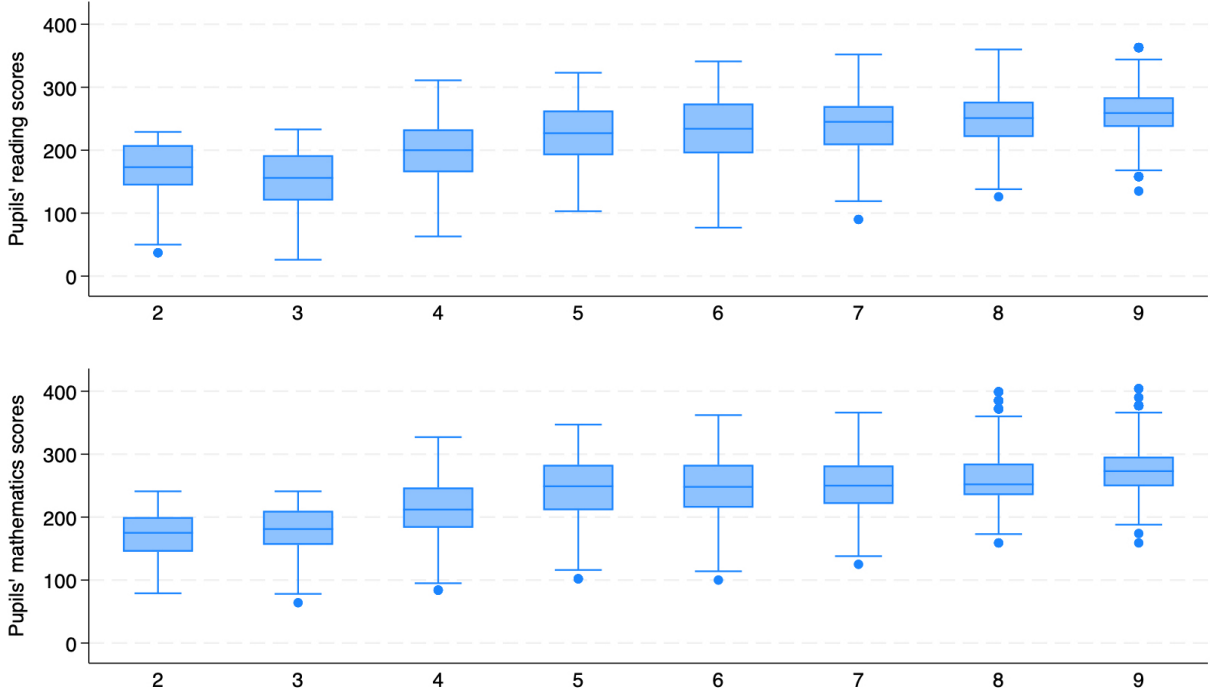
F. Additional Information on Other Relevant Variables

Table A4: Students' Educational Performance Missing Data in 2022 and 2023

Students performance	Before adjustment		After adjustment	
	Missing	Percent	Missing	Percent
Reading - 04/22	2,650	8.34 %	2,650	8.34 %
Mathematics - 04/22	2,689	8.47 %	2,689	8.47 %
Reading - 06/22	2,950	9.29 %	844	2.66 %
Mathematics - 06/22	3,111	9.80 %	881	2.77 %
Reading - 04/22	3,008	9.47 %	511	1.61 %
Mathematics - 09/22	3,060	9.63 %	519	1.63 %
Reading - 09/22	3,001	9.45 %	399	1.26 %
Mathematics - 12/22	3,049	9.60 %	398	1.25 %
Reading - 04/23	5,289	16.65 %	368	1.16 %
Mathematics - 04/23	5,311	16.72 %	375	1.18 %
Reading - 06/23	3,849	12.12 %	351	1.11 %
Mathematics - 06/23	3,973	12.51 %	362	1.14 %
Reading - 09/23	4,006	12.61 %	350	1.10 %
Mathematics -09/23	4,067	12.81 %	356	1.12 %
Reading - 12/23	3,878	12.21 %	348	1.10 %
Mathematics - 12/23	3,985	12.55 %	355	1.12 %

Notes: The table shows the students' education performance missing data in reading and mathematics across eight different dates of application of the Rio Assessments. Our sample includes 31,760 fixed students. The columns below "Before adjustment" show information without any change to fill in missing scores. The columns under "After adjustment" present information after we fill in missing scores with standardised scores from previous Rio Assessments tests. For instance, if a student did not have the standardised reading score from the December 2023 Rio Assessments, we use the standardised reading score from the test held in September 2023 to replace the December 2023 missing score. If the September 2023 reading score was also missing, we used the standardised score from the Rio Assessments held in June 2023. We continue this process until April 2022 if necessary. This is a conservative approach since we use data from when a student was less exposed to the programme.

Figure A5: Control Students' Scores by Grade (Boxplots) in the December 2023 Rio Assessments



Notes: The figure shows the control student score distribution by grade for reading and mathematics in the December 2023 Rio Assessments. Box plots represent the distribution of scores for each grade.

G. Additional information on the Pre-Treatment Summary Statistics

There is a debate among human behaviour scientists regarding the reporting of significance tests for baseline variables in experimental designs. Some argue, including the Consolidated Standards of Reporting Trials (CONSORT), that balance tests are meaningless and even misleading and, therefore, should not be reported (e.g. Senn 1994; Hayes and Moulton 2017; Mutz, Pemantle, and Pham 2019). They claim that using pretreatment summary statistics to validate the hypothesis that the conditions were randomly assigned is meaningless if one already knows the conditions were randomly assigned. Instead, they recommend focusing on the randomisation procedure and attrition problems. Additionally, Mutz, Pemantle, and Pham (2019) discussed the lack of a solid basis for dealing with significant statistical imbalance, concluding that it is incorrect to change a model to include imbalanced covariates unless this variable was previously included in the model because it is supposed to be a predictor of the outcome.

We agree that reporting significance tests for pretreatment variables can be meaningless and misleading when one has a random design study. However, as some researchers still use this approach, as discussed in Gerber et al. (2014) and Gerber et al. (2015), we show SE from t-tests

without any advice or conclusions based on them.³⁷ We aim to avoid misunderstandings as any statistically significant results from the inferential balance tests in randomised experiments are due to chance imbalance, a type I error. Furthermore, we will not use the significance tests to change our model or draw conclusions, following Mutz, Pemantle, and Pham (2019). Readers who may doubt this study’s random sampling or assignment procedures can review the data and codes.

The data in Table A5 comes from the SMERJ. This table presents information on the characteristics and educational outcomes of grade 1–8 students enrolled (during November/December 2021) in the 991 schools under the SMERJ for the 2022 school year. During the 2021 school year, there were 992 grade 1–9 schools under the responsibility of the city of Rio de Janeiro. However, one of the schools was deactivated at the end of 2021. Thus, 991 schools were under the city’s responsibility for the 2022 school year. The data was collected at the end of 2021.

The “Sample” column from Table A5 presents the difference between the sample and population averages with SE of the differences in parentheses. Robust SE are clustered at the school level. Each average, difference between averages, and SE is determined through a regression analysis where the variables in each table row are regressed on a population-sample dummy. This dummy switches off for a population’s student and on for a sample’s student. Sample students have two observations in the data, each with a different identification—one each for the population and sample. The educational outcomes come from standardised reading and mathematics tests, which are meant to be taken by all students across two different dates: April and June 2021.³⁸ We call these tests the Rio Assessments. The students’ scores from the 2021 Rio Assessments were standardised by subject, grade, and test application date using the population as the reference.

Table A5 reveals that the Rio de Janeiro schools have 34.6%, 53.3%, and 11.9% of White, Brown, and Black students, respectively. Additionally, 29.7% of students’ families received support from the Brazilian conditional cash transfer programme, *Bolsa Família*. The differences between the population and the experimental sample students’ averages regarding race, gender,

³⁷Overall F-tests were compromised in our pretreatment data because of missing data. For example, the regression of the population-sample dummy on all the variables from Table A5 had only $\frac{1}{3}$ of the intended observations.

³⁸we do not include any students’ educational outcome information from tests administered in September, October, November, or December 2021 for the pre-treatment analysis since the random assignment process was conducted in September 2021.

and Bolsa Familia are almost null in Table A5. The differences between the population and sample averages regarding students' mathematics and reading scores are also minimal. The population average scores were zero because the population was the reference group for the standardisation procedure.

Table A5: Pre-Treatment Statistics: Student Population

	Valid obs	Population (mean)	Sample (difference/SE)
A. Characteristics			
White (%)	357,623	0.346	-0.003 (0.011)
Brown (%)	357,623	0.533	0.002 (0.008)
Black (%)	357,623	0.119	0.000 (0.005)
Female (%)	390,018	0.486	0.000 (0.003)
Bolsa família (% receiving)	390,018	0.297	0.013 (0.021)
B. Educational outcomes			
Reading - 04/2021	281,002	0.000	0.053 (0.027)
Mathematics - 04/2021	280,546	0.000	0.065 (0.026)
Reading - 06/2021	291,466	0.000	0.035 (0.028)
Mathematics - 06/2021	290,602	0.000	0.040 (0.029)
Students total		393,871	31,760

Notes: This table presents information on the characteristics and educational outcomes of students enrolled (during November/December 2021) in the Rio de Janeiro population of 991 grade 1–8 schools for the 2022 school year. The column “Valid obs” shows the number of observations or students with data available. The “Population” column presents the students’ population averages. The “Sample” column shows the difference between the sample and population averages with standard errors of the differences in parentheses. SE are robust standard errors. Each average, difference between averages, and SE comes from a regression analysis where relevant variables in the table are regressed on a population-sample dummy. This dummy switches off for a population’s student and on for a sample’s student. Sample students have two observations in the data, each with a different identification - one for the population and one for the sample. Students’ scores were standardised by subject, grade, and test application date using the entire population as the reference.

Table A6 reports the average characteristics of principals, pedagogical coordinators, teachers, and schools for the 2022 school year. Data were collected by the SMERJ at the end of 2021. The column “Valid obs” shows the number of observations or schools with data available. The “Sample” column shows the difference between the sample averages and the population. “SE”

refer to robust SE and is reported in parentheses. Each average, difference between averages, and SE comes from a regression (at the school level) in which we regress each relevant variable in the table on a population-sample dummy that switches off for a population's school and on for a sample's school. (Each sample school has two observations in the data with different identifications: one each for the population and the sample.)

Table A6: Pre-Treatment Statistics - School Population

	Valid obs	Population (mean)	Sample (difference/SE)
Principals			
Female (%)	986	0.843	0.007 (0.042)
Age (mean)	986	50.6	0.2 (1.0)
College degree (%)	985	0.812	-0.075 (0.051)
Pedagogical Coordinators			
Female (%)	940	0.893	0.041 (0.031)
Age (mean)	940	44.1	-0.5 (0.9)
College degree (%)	933	0.874	-0.049 (0.046)
Teachers (avg by school)			
Female (%)	991	0.801	0.012 (0.017)
College degree (%)	991	0.842	-0.008 (0.017)
School characteristics			
Students per school (mean)	991	397	0 (23)
Segment I schools (%)	991	0.709	-0.009 (0.053)
Segment II schools (%)	991	0.291	0.009 (0.053)

Notes: This table reports the average characteristics of principals, pedagogical coordinators, teachers, and schools for the 2022 school year. The Municipal Secretary of Education of Rio de Janeiro (SMERJ) collected the data at the end of 2021. The column “Valid obs” shows the number of observations or schools with data available. The “Population” column shows the averages of the Rio de Janeiro population of schools. The ‘Sample’ column shows the difference between the averages of the sample and the population, with the SE of the difference in parentheses. SE are robust standard errors. Each average, difference between averages, and SE comes from a regression (at the school level) in which we regress each relevant variable in the table on a population-sample dummy that switches off for a population's school and on for a sample's school (sample schools have two observations in the data with different identifications, one for the population and one for the sample).

The data presented in Table A6 indicate that a much higher percentage of women work as principals (84.3%), pedagogical coordinators (89.3%), and teachers (80.1%) in the Rio de Janeiro schools than men. Additionally, more than 80% of the professionals in these three careers have a college degree. On average, principals are approximately six years older than pedagogical coordinators. An average of 397 students are observed in grades 1–8 in Rio de Janeiro schools. Segment I schools (grades 1–5) comprise 70.9% of the schools’ population and 70% of the schools’ sample. Segment II schools account for 29.1% of the population and 30% of the sample. The differences between the averages of the population of schools and the experimental sample of schools are almost null.

As expected from a random sampling procedure, Tables A5 and A6 reveal that our experimental sample is very similar, on average, to the Rio de Janeiro population of grade 1–8 students and schools.

H. Regression Adjustment

To check for robustness, we also identify the causal effects of the treatment on school management using a full regression adjustment model with a centred covariate and its interaction with the treatment dummy (Lin 2013; Imbens and Rubin 2015; Negi and Wooldridge 2021; Sloczynski 2022). Let X_s be the school management average score from the management survey conducted in 2021. Thus, the ATE, λ_{ATE} , can be identified for each subject using the following weighted full regression adjustment causal equation:

$$M_s\sqrt{n_s} = \iota\sqrt{n_s} + \lambda_{ATE}Z_s\sqrt{n_s} + (X_s - \bar{X})\gamma\sqrt{n_s} + Z_s(X_s - \bar{X})\delta\sqrt{n_s} + \varepsilon_s\sqrt{n_s}. \quad (4)$$

As in the unadjusted Equation (1), the adjusted Equation (4) identifies the average treatment effect (ATE) of the programme on the schools’ management. SE are robust SE (Chaisemartin and Ramirez-Cuellar 2024). Equation (4) can be used to identify the ATE of the programme on a specific management practice or even a group of management practices.

For robustness, we identify the programme’s ATE on student learning using a full regression adjustment model with centred pretreatment covariates and their interactions with the treatment dummy (Lin 2013; Imbens and Rubin 2015; Negi and Wooldridge 2021; Sloczynski 2022). Let $\mathbf{X}_{isp}$ be student performance in previous Rio Assessments applied in April and June 2021.

Let $\bar{\mathbf{X}}$ be the average of $\mathbf{X}_{isp}$. Therefore, the ATE, β_{ATE} , can be identified for each subject using the following causal equation:

$$Y_{isp} = \alpha + \beta_{ATE}Z_{isp} + (\mathbf{X}_{isp} - \bar{\mathbf{X}}) \cdot \boldsymbol{\gamma} + Z_{isp} \cdot (\mathbf{X}_{isp} - \bar{\mathbf{X}}) \cdot \boldsymbol{\delta} + \epsilon_{isp}. \quad (5)$$

As in the unadjusted Equation 2, the ATE identification and estimation in adjusted Equation 5 relies on using the clustered robust SE with clusters defined at the school pair level along with the noninclusion of pair fixed effects in the regression (Chaisemartin and Ramirez-Cuellar 2024).

For the robustness check, we also identify the causal effect of the intensity of management following the treatment using a full regression adjustment model with centred pretreatment covariates and their interactions with the treatment dummy (Lin 2013; Imbens and Rubin 2015; Negi and Wooldridge 2021; Sloczynski 2022). Let $\mathbf{X}_{isp}$ be the set of pretreatment covariates: student performance in previous Rio Assessments applied in April and June 2021 and the school average score of the 23 management practices in our pretreatment survey. Let $\bar{\mathbf{X}}$ be the average of $\mathbf{X}_{isp}$. Thus, the impact of treatment intensity (school management changes driven by the random assignment/the programme) using a full interaction adjustment model can be identified from the following 2SLS procedure using the random assignment to the treatment as the IV:

$$Y_{isp} = \alpha + \beta_{IV}\hat{M}_{isp} + (\mathbf{X}_{isp} - \bar{\mathbf{X}}) \cdot \boldsymbol{\gamma} + Z_{isp} \cdot (\mathbf{X}_{isp} - \bar{\mathbf{X}}) \cdot \boldsymbol{\delta} + \epsilon_{isp}, \quad (6)$$

where the $\hat{M}_{isp}$ is the fitted M_{isp} from the following first-stage equation:

$$M_{isp} = \omega + vZ_{isp} + (\mathbf{X}_{isp} - \bar{\mathbf{X}}) \cdot \boldsymbol{\nu} + Z_{isp} \cdot (\mathbf{X}_{isp} - \bar{\mathbf{X}}) \cdot \boldsymbol{\theta} + \xi_{isp}.$$

SE are clustered robust SE, with clusters set to be the school pairs used for the random assignment and the noninclusion of pair fixed effects in the regression (Chaisemartin and Ramirez-Cuellar 2024).

I. Additional Information about the Management Groups of Practices' Impact

Table III gives more information regarding the treatment's impact across the five management practice groups. Column (1) reports ATE estimates from weighted regressions of the

2023 group of the practices’ average scores on a treatment dummy without baseline covariates. Column (2) shows ATE estimates generated from weighted regressions of the 2023 group of practices’ average scores on a treatment dummy, the centred average score reached by the schools in the same group of practices in 2021, and the interaction between the centred baseline covariate and the treatment dummy. The weights used are the number of students in grades 1–8 in each school, as discussed in Section III.E. SE are reported in parentheses.

In the unadjusted regression, Table III reports that the treatment’s impact was 1.530 SD for the target-setting group of practices and 1.069 SD for the leadership group. Both estimates are statistically significant at the 1% level. Although the programme’s causal effect on the operation group was 0.370 SD in the unadjusted regression, it is not statistically significant. However, the ATE estimate of the treatment’s impact on the operation group reaches 0.400 SD in the adjusted regression and is statistically significant at the 10% level. The ATE estimate for the monitoring group of practices was 0.324 SDs; however, it is not statistically significant. People management did not change significantly as a result of the treatment.

J. Additional Information on Robustness Checks

Table A7 shows the causal effects of the treatment on student learning under different specifications. All outcome variables were generated in December of 2023, two years after the programme started to be implemented. The sample includes all students (31,760) enrolled in grades 1 to 8 at the beginning of 2022 in one of the 80 sample schools randomly selected from the Rio de Janeiro school population. These students registered to a sample school in the regular enrollment period—November and December 2021—before the experiment began. The students’ reading and mathematics scores from the December 2023 Rio Assessment represent their learning. The scores were standardised to have a mean of 0 and an SD of 1 in each grade and subject regarding the control group. “Mgmt ([1,5])” represents the schools’ average management score from 1 to 5. “Mgmt ([1,5])” comes from our survey conducted in December 2023 that adopted the WMS methodology.

Table A7 shows results from analyses with 31,710 and 31,706 valid student scores for reading and mathematics, respectively. Columns (1) and (3) report (ATE) estimates from regressions without and with baseline covariates, respectively. Columns (2) and (4) show IV estimates generated from regressions without and with baseline covariates, respectively. IV estimates can

be interpreted as the causal effects of a one-point change in school management, driven by the treatment, on students' educational outcomes.

Table A7: The Treatment's Causal Effects - Alternative Students' Scores (CTT)

	No covariates		With covariates	
	(1)	(2)	(3)	(4)
	ATE	Mgmt ([1,5])	ATE	Mgmt ([1,5])
	OLS	IV	RA	IV/RA
Reading (SD)	0.209*** (0.055)	0.630*** (0.252)	0.174*** (0.042)	0.522*** (0.192)
Mathematics (SD)	0.220*** (0.053)	0.662*** (0.247)	0.194*** (0.044)	0.583*** (0.201)

Note: The table shows the impact of the treatment, SMEP, under different specifications. All outcome variables are from December 2023, two years after the beginning of the treatment. The sample includes all students (31,760) enrolled in grades 1 to 8 at the beginning of 2022 in one of the 80 randomly selected schools from the Rio de Janeiro population of schools. Before the experiment began, these students registered to a sample school in the regular enrollment period in November and December 2021. The students' reading and mathematics scores from the December 2023 Rio Assessment represent their learning. The scores were standardised to have a mean of zero and a standard deviation of one for each grade and subject in relation to the control group. "Mgmt ([1,5])" represents the schools' average management score from one, worst management, to five, the best management. "Mgmt ([1,5])" represents the schools' average management score that goes from 1, worst management, to 5, the best management. "Mgmt ([1,5])" is derived from a survey conducted in December 2023 that adopted the World Management Survey (WMS) methodology. From 31,760 sample students, there are 31,710 and 31,706 valid students' scores for reading and mathematics, respectively. Columns (1) and (3) report average treatment effects (ATE) estimates from regressions without and with baseline covariates, respectively. Columns (2) and (4) show IV estimates generated from regressions without and with baseline covariates, respectively. IV estimates can be interpreted as the causal effects of a 1-point change (on a 1–5 scale) in school management, driven by the random assignment/treatment, on students' educational outcomes. Columns (3) and (4) report ATE and IV estimates from regressions with the following centred baseline covariates (and their interaction with treatment): students' reading and mathematics scores from previous Rio Assessments that were applied in April and June 2021, and schools' average management score from our 2021 management survey. Clustered robust standard errors are reported in parentheses. The clusters for these standard errors are the school pairs used for the random assignment procedure. Significance at the 1%, 5%, and 10% levels indicated by ***, **, and *, respectively.

Columns (3) and (4) from Table A7 report ATE and IV estimates from regressions with the following centred baseline covariates (and their interaction with treatment): students' reading and mathematics scores from previous Rio Assessments applied in April and June 2021 and schools' average management score from our 2021 management survey. The results based on the IRT scores from the Rio Assessments reported in Table IV are very similar to the estimates reported in Table A7 based on the CTT scores from the same Rio Assessments.

Table A8: Sharpened False Discovery Rate (FDR) q-values

Location	Identifier	p-values	bky06_qval
Table IX - Unadjusted	Reading (ATE/OLS)	0.000	0.002
	Maths (ATE/OLS)	0.000	0.002
	Management (ATE/OLS)	0.001	0.003
	Reading (IV)	0.005	0.009
	Maths (IV)	0.007	0.011
Table IX - Adjusted	Reading (ATE/RA)	0.000	0.002
	Maths (ATE/RA)	0.000	0.002
	Management (ATE/RA)	0.001	0.002
	Reading (IV/RA)	0.002	0.005
	Maths (IV/RA)	0.003	0.007
IV.A Absence	Absence 22	0.061	0.047
	Absence 23	0.000	0.002
Table X - Unadjusted	Target 23 (ATE/OLS)	0.000	0.002
	Leadership 23 (ATE/OLS)	0.000	0.002
Table X - Adjusted	Target 23 (ATE/RA)	0.000	0.002
	Leadership 22 (ATE/RA)	0.000	0.002
Figure XIV	Planning standardisation (o1)	0.828	0.43
	Instruction personalisation (o2)	0.362	0.248
	Data-driven planning (o3)	0.153	0.109
	Adopting best practices (o4)	0.030	0.027
	Continuous improvement (m5)	0.588	0.356
	Performance tracking (m6)	0.028	0.026
	Performance review (m7)	0.189	0.127
	Performance dialogue (m8)	0.717	0.407
	Consequence management (m9)	0.928	0.46
	Target balance (t10)	0.000	0.001
	Target interconnection (t11)	0.000	0.002
	Target time horizon (t12)	0.001	0.003
	Target stretch (t13)	0.022	0.021
	Target clarity/comparability (t14)	0.008	0.012
	Rewarding high performers (p15)	0.933	0.46
	Fixing poor performers (p16)	0.520	0.327
	Promoting high performers (p17)	1.000	0.491
	Managing talent (p18)	0.201	0.134
	Retaining talent (p19)	0.409	0.28
	Attracting employees (p20)	0.228	0.151

Table A8: Sharpened False Discovery Rate (FDR) q-values

Location	Identifier	p-values	bky06_qval
Figure XV	Leadership vision (l21)	0.000	0.002
	Leadership accountability (l22)	0.034	0.029
	Clearly defined roles (l23)	0.150	0.109
	Apr-22	0.013	0.016
	Jun-22	0.017	0.019
	Sep-22	0.021	0.021
	Dec-22	0.015	0.017
	Apr-23	0.001	0.004
	Jun-23	0.001	0.002
	Sep-23	0.000	0.002
	Dec-23	0.000	0.002
IV.C Heterogeneous	Management 21	0.004	0.007
	Segments (I or II)	0.768	0.421
	Grade 2	0.743	0.419
	Grade 3	0.756	0.42
	Grade 4	0.433	0.289
	Grade 5	0.973	0.48
	Grade 6	0.572	0.351
	Grade 7	0.845	0.431
	Grade 8	0.923	0.46
	School size	0.425	0.288
	Female	0.865	0.437
	Brown	0.304	0.207
	Black	0.569	0.351
	Asian/Indigenous	0.526	0.327
	Bolsa Familia	0.832	0.43
	Quartil 2	0.365	0.248
	Quartil 3	0.179	0.122
	Quartil 4	0.657	0.387
Table XI - Unadjusted	Reading (ATE/OLS)	0.001	0.002
	Maths (ATE/OLS)	0.000	0.002
	Reading (IV)	0.012	0.016
	Maths (IV)	0.007	0.011
Table XI - Adjusted	Reading (ATE/RA)	0.000	0.002
	Maths (ATE/RA)	0.000	0.002
	Reading (IV/RA)	0.007	0.011

Table A8: Sharpened False Discovery Rate (FDR) q-values

Location	Identifier	p-values	bky06_qval
	Maths (IV/RA)	0.004	0.007
IV.E Robustness II	Reading (ATE/RA)	0.001	0.002
	Maths (ATE/RA)	0.000	0.002
	Reading (IV/RA)	0.008	0.012
	Maths (IV/RA)	0.009	0.012
V - CBA/CEA	Learning (ATE/OLS)	0.000	0.002
	Learning (IV)	0.005	0.009

Notes: This table presents information on the p-values generated by this study's analyses and their corresponding sharpened false discovery rate (FDR) q-values. The FDR is the expected proportion of type I errors (false rejections) (Anderson 2008).

K. Cost-Benefit Analysis - Krueger Approach

We follow Krueger (2003) to calculate the internal rates of return (IRRs) of the SMEP based on the expected income benefits from the increased student educational performance. For comparison, let us assume parameter values following Fryer (2017a). We consider students entering the 8th grade at the beginning of 2022 without loss of generality. Suppose that the earnings of the current labor force in the United States represent the profile of earnings by age that the average student who entered the 8th grade (13 years old) in 2022 will experience when they enter the labor market. For comparison, we use our impact estimates and costs, adjusted by PPP, in the U.S. context.

Denote the annual cost per student of the programme in year t as C_t . We used the ingredients method to calculate the SMEP cost (Dhaliwal et al. 2013). Details about the programme's costs are provided in Section M in the Appendix. The program lasted two years. Let E_t be the individual's annual earnings each year t from age 18 to retirement at age 65. δ represents the increase in earnings associated with a 1 SD increase in either mathematics or reading. According to Neal and Johnson (1996), Krueger (2003), and Chetty, Friedman, and Rockoff (2014), the increase in earnings associated with an increase of 1 SD in reading or mathematics is between 8% and 20%. We used $\delta = 0.12$ to make the comparison with Fryer (2017a) possible.

Let g be the real earnings growth rate by year. Actual earnings and productivity have historically grown between 1 and 2 % (Krueger 2003). To make the comparison with Fryer

(2017a) possible, let $g = 0.02$. Lastly, let β_{ATE_m} and β_{ATE_r} be the programme’s impact on students’ mathematics and reading test scores due to being assigned to a treatment school. The IRR, r , is calculated by determining the discount rate at which the present value of future cash flows (benefits) equals the initial investment (programme cost).

$$\sum_{t=1}^2 C_t/(1+r)^t = \sum_{t=18}^{65} E_t \times (1+g)^{t-13} \times \delta \times (\beta_{ATE_m} + \beta_{ATE_r})/(1+r)^{t-13} \quad (7)$$

The superscripts on $(1+g)$ and $(1+r)$ are $t-13$ as the students starting the 8th grade are 13 years old, five years from being 18 years old (and entering the labor market). We also calculate the IRR considering the treatment intensity analysis (IV) by replacing the impact parameters from the equation above with the IV impact estimates.

Table A9 reports the annual cost per student considering treatment and control arms, student gains in reading and mathematics, and the previous IRR for the interventions discussed. Apart from our experiment and the Jovem de Futuro, all the other studies’ values in Table A9 come from Fryer (2017a,b). We updated all cost values from Fryer (2017a) to January 2022. The SMEP and the Jovem de Futuro costs are in USD PPP GDP (2022). The column ‘Gains’ shows the sum of the causal effect of each intervention on student reading and mathematics in SD.

The different interventions in education presented in Figure X are related to different areas.

- **Management:** our “SMEP ATE,” and our “SMEP High Implementation (IV)” ; Fryer (2017a) “Houston High Implementation (HI),” “Houston Principal Staying (PS),” and “Houston Overall”; Barros et al. (2019) and Barros et al. (2021) “Jovem de Futuro.”
- **Early Childhood:** Puma et al. (2010) ‘Head Start Impact Study’.
- **Charter schools:** Fryer (2014) ‘Injecting Best Practices Elementary Schools (ES)’, and ‘Injecting Best Practices Secondary Schools (SE)’; Dobbie and Fryer (2011) ‘Harlem Children’s Zone Elementary Schools (ES)’, and ‘Harlem Children’s Zone Middle Schools (ME)’; Curto and Fryer (2014) ‘SEED Schools’.
- **Teacher Incentives:** Glazerman et al. (2013) ‘Talent Transfer Initiative’.
- **Teacher Certification:** Glazerman, Mayer, and Decker (2006) ‘Teach For America’.

Table A9: Cost-Benefit Analysis (CBA)

	Cost per student per year (treatment)	Cost per student per year (control)	Gains Read+Math (SD)	Internal rates of return (IRR)
Management				
SMEP, High Implementation (IV)	\$15.22	\$0.00	1.39	227%
SMEP (ATE)	\$15.22	\$0.00	0.46	165%
Houston Overall	\$11.55	\$0.42	0.06	79%
Houston HI	\$11.55	\$0.42	0.12	96%
Houston PS	\$11.55	\$0.42	0.11	94%
Jovem de Futuro	\$216.79	\$0.00	0.20	48%
Early Childhood				
Head Start Impact Study	\$12,002.31	\$3,782.33	0.32	9 %
Charter Schools				
Injecting Best Practices ES	\$433.60	\$ 0.00	0.26	35 %
Injecting Best Practices SS	\$2,243.74	\$ 0.00	0.13	18 %
Harlem Children's Zone ES	\$25,560.18	\$16,502.98	0.31	11 %
Harlem Children's Zone MS	\$25,560.19	\$16,502.98	0.28	12 %
SEED Schools	\$51,904.45	\$27,122.53	0.42	9 %
Teacher Incentives				
Talent Transfer Initiative	\$788.67	\$0.00	0.24	28 %
Teacher Certification				
Teach For America	\$4,455.20	\$0.00	0.18	12 %
Class Size				
Tennessee STAR	\$6,111.44	\$0.00	0.24	10 %
Professional Dev				
Success for All	\$1,024.49	\$0.00	0.09	14 %
Tutoring				
Experience Corps	\$1,045.16	\$0.00	0.08	14 %
Curriculum				
Enhanced Reading	\$2,519.73	\$0.00	0.18	22 %
Financial Incentives				
Coshocton Incentive Program	\$91.62	\$0.00	0.12	49 %
New York, Dallas, Chicago	\$429.01	\$0.00	0.00	-

Notes: This table presents the per-student costs per year, the impact and the internal rates of return for our experiment, 17 interventions presented in Fryer (2017a) and the Jovem de Futuro programme analysed by Barros et al. (2019) and Barros et al. (2021). Apart from our experiment and the Jovem de Futuro Programme, all other studies' values come from Fryer (2017a,b). We updated all the cost values reported by Fryer (2017a) to January 2022. The IRR is the discount rate that equates the cost of an intervention with the present value of the expected future earnings inflows that the intervention generates. We calculated the IRR for our Science and Management for Education Programme (SMEP) and the Jovem de Futuro following the methodology developed by Krueger (2003). The Science and Management for Education Programme and the Jovem de Futuro costs are in USD PPP GDP (2022).

- **Class Size:** Krueger (1999) ‘Tennessee STAR’.
- **Managed Professional Development:** Borman, Slavin, and Cheung (2007) ‘Success for All’.
- **Tutoring:** Morrow-Howell et al. (2009) ‘Experience Corps’.
- **Curriculum:** Somers et al. (2010) ‘Enhanced Reading Opportunities’.
- **Financial Incentives:** Bettinger (2012) ‘Coshocton Incentive Programme’; and Fryer (2011) ‘New York, Dallas, Chicago’.

L. Cost-Effectiveness Analysis (J-PAL Approach)

We also used the J-PAL approach to measure the cost-effectiveness of our programme and make comparisons with other interventions (Dhaliwal et al. 2013; J-PAL 2020). We followed the sequence recommended by J-PAL to address inflation, exchange rates, and the present value of the programme costs. The program’s local cost data in Brazilian Reais (R\$) is translated into US dollars using the exchange rate from the year the costs were incurred, 2022 and 2023. The 2023 costs are deflated back to the actual value in 2022 prices using the average annual U.S. inflation rate. Then, the present value of the costs incurred in 2023 deflated to 2022 is calculated using a 10% discount rate. Following this, we used the average U.S. inflation rate to deflate the total costs from 2022 to 2011. Lastly, the costs in 2011 US dollars were translated into 2011 USD (PPP GDP).

We ran the same ATE and treatment intensity regressions (without covariates) shown in Section III.E; however, we used the average of student reading and mathematics standardised scores as the outcome variable. The ATE and treatment intensity (IV) estimates using the subjects’ average as the outcome are called the “average impact” of the treatment for this cost-effectiveness analysis.

As the average impact was measured two years after the beginning of the treatment, a 10% discount rate should be applied, as recommended by Dhaliwal et al. (2013). The total impact of the treatment is the average impact in SD multiplied by the number of students in the treated schools, considering the ATE estimates. Regarding the intensity treatment estimates (IV), we considered the number of students in high-implementation schools. We run a regression of the

overall management of the schools on a treatment dummy for each of the 40 pairs used for the pair-matching random assignment. The goal is to identify high-implementation schools. For practical purposes, we define a high-implementation school as a treated school that changed its management level to equal or above 0.5 points compared with the control school within its pair. There are 19 high-implementation schools. This group of 19 schools changed their management on average by 1 point due to the treatment compared with the control schools within their pairs.

Finally, the total cost (ATE) in 2011 US dollars (PPP GDP) divided by the total impact (ATE) in SD is the cost per additional SD. Thus, if we divide 100 by the cost per additional SD, we have the additional SD per \$100.00, which is the cost-effectiveness indicator recommended by the J-PAL. We do the same for the IV estimates of the high-implementation schools using only their costs (19 schools).

Differently from Dhaliwal et al. (2013) and J-PAL (2020) that used a 10% of significance level cut-off to select experiments to have their cost-effectiveness calculated, we only show in Table A10 the “Additional SD per \$100” for the randomised experiments that are significant at the 5% level. Apart from the nonsignificant interventions, the statistically significant programmes at the 5% level are organised in descending order of their cost-effectiveness measure: “Additional SD per \$100.” In Table A10, “Average Impact (SD)” represents a programme’s average impact on students’ test scores across subjects (if more than one subject was used) in SD. All costs are in 2011 USD (PPP GDP).

The 30 interventions by country are presented in Table A10:

- **Kenya:** Duflo, Dupas, and Kremer (2011) “Streaming by achievement,” “Extra contract teacher + streaming,” and “Contract teachers”; Glewwe, Kremer, and Moulin (2009) “Textbooks for the top quintile,” “Textbooks,” and “Flipcharts”; Kremer, Miguel, and Thornton (2009) “Girls’ merit scholarships”; Duflo, Dupas, and Kremer (2015) “Reducing class size”; Glewwe, Ilias, and Kremer (2010) “Teacher incentives (year 1),” “Teacher incentives (year 2),” and “Teacher incentives (long-run).”
- **Indonesia:** Pradhan et al. (2014) “Linking school to local govt,” “Electing school & linking to local govt,” “School committee grants,” and “Training school committees.”
- **India:** Banerjee et al. (2007) “Remedial education”; Banerjee et al. (2007) “Individually-paced computer assisted learning”; Banerjee et al. (2007) “Reducing class size”; Borkum,

He, and Linden (2012) “Building/improving libraries”; Muralidharan and Sundararaman (2010) “Diagnostic feedback”; and Duflo, Hanna, and Ryan (2012) “Camera monitoring.”

- **Afghanistan:** Burde and Linden (2013) “Village-based schools.”
- **Philippines:** Abeberese, Kumler, and Linden (2014) “Read-a-thon.”
- **Malawi:** Baird, McIntosh, and Özler (2011) “Unconditional cash transfers” and “Minimum cond cash transfers.”
- **Madagascar:** Lassibille et al. (2010) “Providing earnings information.”
- **Gambia:** Blimpo and Evans (2013) “School committee grants” and “Grants & training for school committee.”
- **Colombia:** Barrera-Osorio and Linden (2009) “Adding computers to classrooms.”
- **Peru:** Cristia et al. (2017) ‘One Laptop Per Child’.

Table A10 shows that the ‘SMEP High Implementation (IV)’ is the fourth most cost-effective intervention among the programmes assessed based on the “Additional SD per \$100 (PPP).” “SMEP High Implementation (IV)” represents the programme’s impact on high-implementation schools (schools that changed their management level by around 1 point on a scale from 1 to 5). In this case, the intervention cost-effectiveness achieves a 4.80 “Additional SD per \$100 (PPP),” showing the power of management to improve students’ learning and school productivity.

Table A10 shows that the three interventions performing better than our high-implementation treatment regarding “Additional SD per \$100 (PPP)” have their 95% confidence intervals’ lower bound technically **at zero SDs (SD)**. While the high implementation of our programme has a 95% lower bound of 0.210 SD, Pradhan et al. (2014) “Linking school to local govt” in Indonesia, Duflo, Dupas, and Kremer (2011) “Streaming by achievement” in Kenya and Pradhan et al. (2014) “Electing school & linking local govt” in Indonesia have 95% confidence intervals’ lower bounds of 0.034 SD and 0.025 SD, respectively. The chance of having a null effect is much greater in their case than the high-implementation level of our intervention. From a public policy perspective, the high implementation of SMEP should rank first because the intervention has a much lower level of uncertainty on nonzero effects.

Table A10: Cost-Effectiveness Analysis (CEA)

Programmes	Average Impact SD	95% Lower Bound	95% Upper Bound	Additional SD per \$100 (PPP)
Linking school to local govt, Indonesia	0.165	0.034	0.296	26.10
Streaming by achievement, Kenya	0.176	0.025	0.327	16.44
Electing school & linking local govt, Indonesia	0.216	0.034	0.398	10.05
SMEP High Implementation (IV), Brazil	0.700	0.210	1.190	4.80
SMEP (ATE), Brazil	0.232	0.122	0.342	1.71
Textbooks for top quintile, Kenya	0.218	0.030	0.406	1.68
Remedial education, India	0.138	0.046	0.230	1.29
Village-based schools, Afghanistan	0.588	0.302	0.874	0.98
Extra contract teacher+streaming, Kenya	0.248	0.068	0.428	0.93
Read-a-thon, Philippines	0.130	0.032	0.228	0.68
Individual computer assisted learning, India	0.475	0.342	0.608	0.65
Contract teachers, Kenya	0.228	0.114	0.342	(0.14)
Unconditional cash transfers, Malawi	-0.030	-0.195	0.135	
Minimum cond cash transfers, Malawi	0.202	-0.029	0.433	
Girls' merit scholarships, Kenya	0.270	-0.044	0.584	N
Providing earnings information, Madagascar	0.202	-0.006	0.410	O
Reducing class size, Kenya	0.074	-0.098	0.246	T
Textbooks, Kenya	0.023	-0.148	0.194	
Flipcharts, Kenya	-0.006	-0.101	0.089	S
Reducing class size, India	0.056	-0.077	0.189	I
Building/improving libraries, India	-0.045	-0.168	0.078	G
School committee grants, Indonesia	0.129	-0.055	0.313	N
School committee grants, Gambia	0.030	-0.146	0.206	I
Computers to classrooms, Colombia	0.109	-0.095	0.313	F
One Laptop Per Child, Peru	0.003	-0.105	0.111	I
Diagnostic feedback, India	0.002	-0.086	0.090	C
Teacher incentives (year 1), Kenya	0.048	-0.072	0.168	A
Teacher incentives (year 2), Kenya	0.136	-0.003	0.275	N
Teacher incentives (long-run), Kenya	0.077	-0.062	0.216	T
Camera monitoring, India	0.170	-0.006	0.346	
Training school committees, Indonesia	-0.049	-0.184	0.086	
Grants/training for school cmte, Gambia	-0.080	-0.256	0.096	

Notes: This table shows the cost-effectiveness of our programme, “SMEP (ATE), Brazil,” the high implementation of our programme, “SMEP High Implementation (IV), Brazil,” and 30 randomised interventions on education analysed by J-PAL (2020). We follow the J-PAL approach presented in Dhaliwal et al. (2013) to calculate the cost-effectiveness of our treatment. However, unlike Dhaliwal et al. (2013), who use a cut-off of 10% to select experiments to have their cost-effectiveness calculated, we only show the ‘Additional SD per \$100’ for the random experiments that are significant at the 5% level. ‘SD’ means standard deviation. Apart from the non-significant interventions, the significant programmes at the 5% level are organised in descending order of their cost-effectiveness measure: “Additional SD per \$100.” “Average Impact (SD)” shows the average impact of a programme on pupils’ test scores in standard deviations (SD) across subjects (if more than one subject was used). All costs are in USD (PPP GDP) 2011.

The “SMEP (ATE)” cost-effectiveness reaches 1.71 “Additional SD per \$100 (PPP).” Table A10 shows the three interventions (apart from the “SMEP, High Implementation (IV),” ranking higher than our treatment regarding “Additional SD per \$100 (PPP)” have their 95% confidence intervals’ lower bound technically **at 0 SD**. While our programme has a 95% lower bound of 0.130 SD, Pradhan et al. (2014) “Linking school to local govt” in Indonesia, Duflo, Dupas, and Kremer (2011) “Streaming by achievement” in Kenya, and Pradhan et al. (2014) “Electing school & linking to local govt” in Indonesia have 95% confidence intervals’ lower bounds of 0.034 SD, 0.025 SD and 0.034 SD, respectively. They have a much larger chance than our treatment (ATE) of having no impact. In terms of policy, the ATE estimates of our treatment would rank second (only behind the “SMEP High Implementation (IV)” because our intervention has a much lower uncertainty regarding the zero impact than the three interventions mentioned.

M. Additional Information on the Programme’s Cost

Table A11: Programme Nominal Costs in 2022 and 2023

2022 Nominal Costs in Brazilian Reais (BRL)			
	Quantity	Cost by Unit	Total by Item
Manager	1	130,000.00	130,000.00
Full time staff	5	97,500.00	487,500.00
Transport	100	100.00	10,000.00
Seminars	2	2,500.00	5,000.00
Total			632,500.00
2023 Nominal Costs in Brazilian Reais (BRL)			
	Quantity	Cost by Unit	Total by Item
Manager	1	130,000.00	130,000.00
Full time staff	5	97,500.00	487,500.00
Transport	100	100.00	10,000.00
Seminars	2	2,500.00	5,000.00
Total			632,500.00

Note: The table shows the programme’s nominal costs by item and year. BRL = Brazilian Reais.

The equivalent of six full-time civil servants (one of them was the manager) worked for two years in the intervention. These professionals were already civil servants of Rio de Janeiro’s

public administration. Their salaries represent the main programme cost.

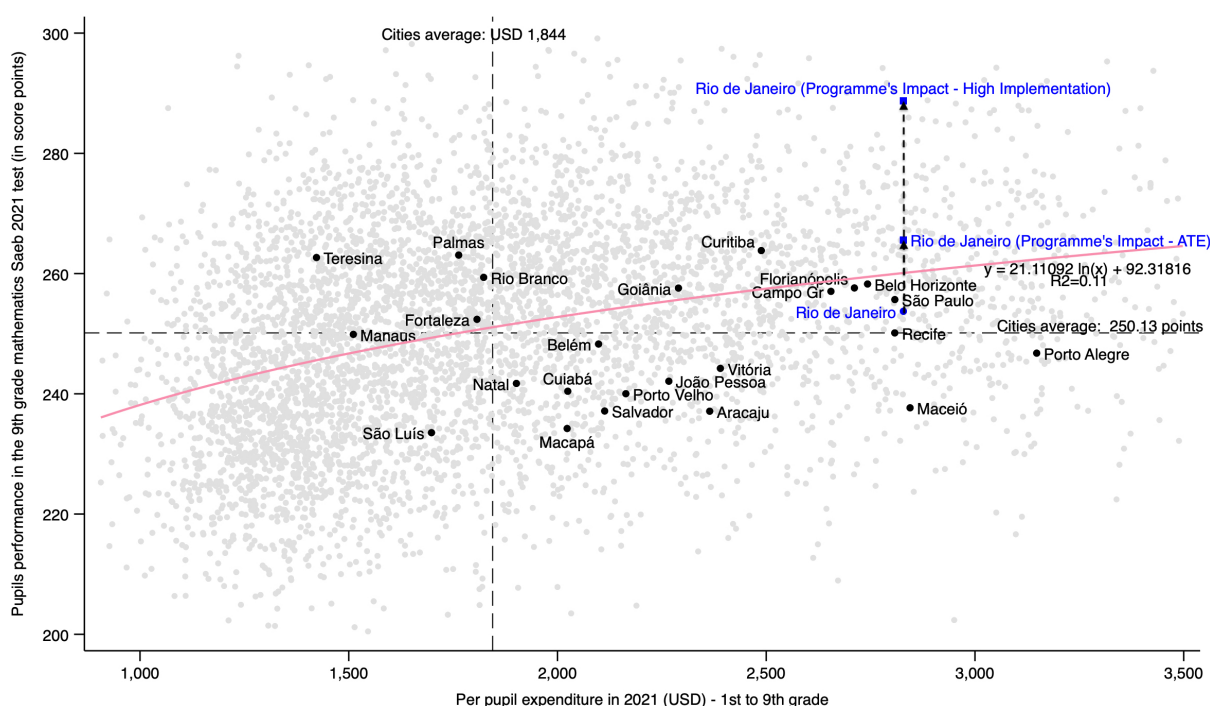
Our study sample comprised 15,626 and 16,134 students in treatment and control schools, respectively. The control schools had no cost. Dividing the 2023 total cost by the Brazilian inflation rate 2022/2023 (1.0635) provides the total cost of 2023 in 2022 values, that is, R\$ 594,734.37. Dividing the total cost of the programme in 2022 and 2023 (in 2022 values), R\$ 1,227,234.37 = 632,500.00 + 594,734.37, by the PPP conversion factor in 2022, 2.58, gives us the total cost of the programme of USD (PPP GDP) 508,956.96.³⁹ Dividing this total cost by 2 (two years of intervention) and the number of treatment students gives us a yearly per-student USD cost of (PPP GDP) 15.22.

N. Additional Information on External Validity

Figure A6 shows the Brazilian cities' average performance in 9th-grade mathematics in the Saeb 2021 versus their public per-student expenditure (grades 1–9) in 2021 in US dollars.⁴⁰ The capital cities of the federated states are highlighted. The data are from Siope and Inep.

Because the Rio Assessments have the same scale and items (TRI) used by Saeb (1 SD is equivalent to 50 points), it is straightforward to translate our impact estimate in mathematics (ATE), 0.237 SD, as 11.85 points in the Saeb scale. The same is done for the high implementation of the program (IV estimates), that is, 0.714 SD of impact, when the management improved by 1 point, can be translated to approximately 35 points in the Saeb scale. The SMEP should significantly affect student learning in Rio de Janeiro.

Figure A6: Programme's Impact in the City of Rio de Janeiro Put in Brazil's Context



Notes: The figure shows Brazilian cities' performance in the ninth-grade mathematics test of the 2021 Saeb compared to the 2021 cities' public expenditure per student from grades 1 to 9 in USD (USD 1.00 = R\$ 5.06 in 19/10/2023, where R\$ means Brazilian Reais). The capital cities of the federated states are highlighted. The data are from Siope and Inep.

We now conduct a thoughtful exercise exploring the expansion of the intervention to all

³⁹OECD Data Explorer

⁴⁰USD 1.00 = R\$ 5.06 in 19/10/2023, where R\$ means Brazilian Reais.

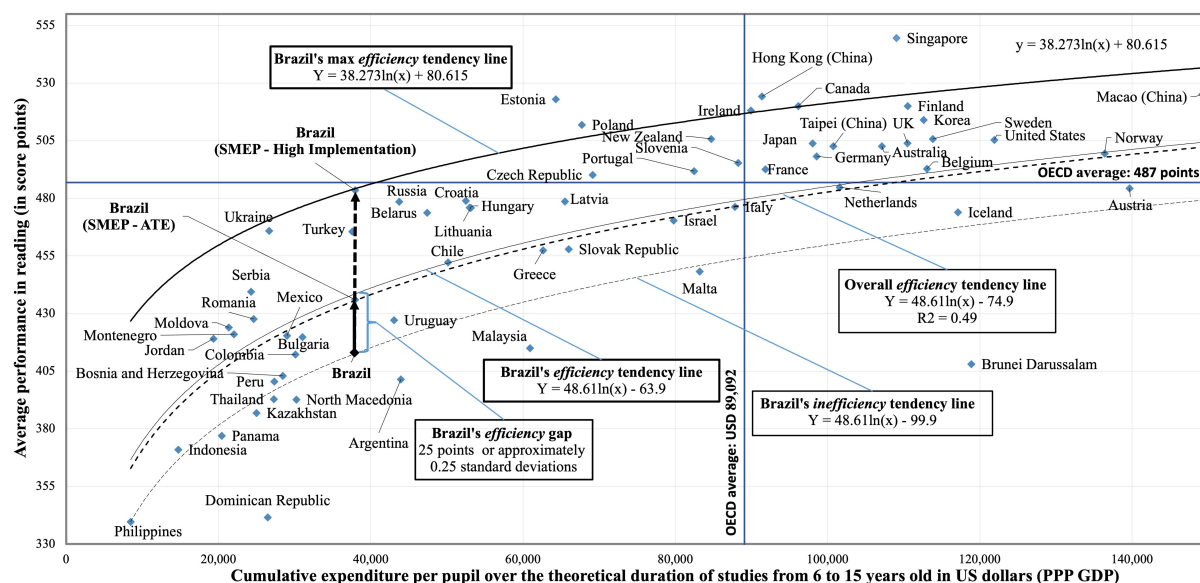
schools in Brazil, assuming a similar cost and impact on educational outcomes. Figure A7 displays the average reading performance (in score points) of students by country in the OECD's PISA 2018⁴¹ versus the cumulative expenditure per student enrolled in primary or secondary education between the ages of 6 and 15 (in USD, corrected by PPP).⁴²

The “Overall *efficiency* tendency line” in Figure A7 is a logarithmic tendency line representing the average relationship between reading performance in PISA 2018 and the cumulative per-student spending on education, considering all countries participating in PISA 2018. We use the word *efficiency* because the overall tendency line can also be understood as the average PISA 2018 countries’ efficiency in using educational resources (overall efficiency), that is, education resources vs. educational performance.

The dashed thin line displays “Brazil’s *inefficiency* tendency line” (actual Brazil results in 2018). Brazil’s efficiency gap in education corresponds to the distance between “Brazil’s *inefficiency* tendency line” and the “Overall *efficiency* tendency line.” This distance shows how less efficient Brazil was compared to the average education spending per student vs. educational performance among all countries participating in PISA 2018.

As 1 SD corresponds to approximately 100 score points in PISA, the programme’s impact (ATE) of 0.23 SD in reading equates to an improvement of approximately 23 PISA score points. “Brazil’s *efficiency* tendency line” illustrates how the SMEP could enhance student performance nationwide if the intervention were implemented at scale with very little change in the x-axis due to its low cost. The SMEP could close the educational efficiency gap between Brazil and the other countries participating in the 2018 PISA.

Figure A7: Programme’s Impact in Brazil in a Global Context



Notes: The figure displays the average reading performance (in score points) of students in the OECD's PISA 2018 by country versus the cumulative expenditure per student enrolled in primary or secondary education between the ages of 6 and 15 (in equivalent USD converted using PPPs for GDP). Data source: PISA 2018 Results (Volume I) - © OECD 2019.

Interestingly, even by doubling annual per-student spending (from USD PPP 37,954 to USD PPP 75,908), Brazil's PISA performance would be below that of Chile (Chile's cumulative per-student expenditure on education is USD PPP 50.149,00). However, adopting the programme

⁴¹PISA measures 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges.

⁴²Luxembourg and Qatar are included in the regressions; however, they are not displayed in the graph as they are outliers in expenditures and would unnecessarily extend the x-axis.

(ATE) could bring Brazil closer to Chile in terms of efficiency, as both countries align on similar efficiency lines.

The final line to be discussed in Figure A7 is the “Brazil max *efficiency* tendency line,” represented by a solid thick black line. This line illustrates how Brazil could have improved its educational performance and efficiency if it had implemented the programme at a high level, enhancing school management by around 1 point on a 1–5 scale. Under this scenario, Brazil’s performance in PISA would have increased by 68 score points, equivalent to the high-level implementation impact of our treatment (IV estimates) on student educational outcomes, corresponding to an increase of 0.68 SD.

As shown in Figure A7, if Brazil had implemented the programme at a high level, improving school management by 1 point, the SMEP would supposedly elevate Brazil’s educational efficiency to a level comparable with developed countries such as Italy, the Netherlands, and Norway. Furthermore, in this scenario, Brazil would have completely closed the educational gap with the OECD average (72 score points on the reading for PISA 2018) while keeping costs almost unchanged, as indicated by the vertical dotted arrow in Figure A7.

REFERENCES

- Abadie, Alberto and Guido W. Imbens, “Bias-Corrected Matching Estimators for Average Treatment Effects”, *Journal of Business & Economic Statistics* 29 2011, 1–11. 10.1198/jbes.2009.07333
- Abdulkadiroğlu, Atila et al., “Accountability and Flexibility in Public Schools: Evidence from Boston’s Charters And Pilots”, *The Quarterly Journal of Economics* 126 May 2011, 699–748. 10.1093/qje/qjr017
- Abeberese, Ama Baafra, Todd J. Kumler, and Leigh L. Linden, “Improving Reading Skills by Encouraging Children to Read in School:” *Journal of Human Resources* 49 2014, 611–633. 10.3368/jhr.49.3.611
- Alan, Sule and Elif Kubilay, “Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting”, *American Economic Review* 115 Feb. 2025, 365–407. 10.1257/aer.20240374
- Anderson, Michael L., “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”, *Journal of the American Statistical Association* 103 2008, 1481–1495. 10.1198/016214508000000841
- Angrist, Joshua D. and Guido W. Imbens, “Two-stage least squares estimation of average causal effects in models with variable treatment intensity”, *Journal of the American Statistical Association* 90 1995, 431–442. 10.1080/01621459.1995.10476535
- Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters, “Explaining Charter School Effectiveness”, *American Economic Journal: Applied Economics* 5 2013, 1–27. 10.1257/app.5.4.1
- Angrist, Joshua D. et al., “Inputs and Impacts in Charter Schools: KIPP Lynn”, *The American Economic Review* 100 2010, 239–243. 10.1257/aer.100.2.239
- “Who Benefits from KIPP?”, *Journal of Policy Analysis and Management* 31 2012, 837–860. 10.1002/pam.21647
- Angrist, Noam et al., “Measuring Human Capital using Global Learning data”, *Nature* 592 2021, 403–408. 10.1038/s41586-021-03323-7
- Angrist, Noam et al., “How To Improve Education Outcomes Most Efficiently? A Review of the Evidence Using a Unified Metric”, *Journal of Development Economics* 172 2025, 103382. 10.1016/j.jdeveco.2024.103382

- Avvisati, Francesco and Pauline Givord, “The Learning Gain Over One School Year Among 15-Year-Olds: An Analysis of Pisa Data for Austria And Scotland (United Kingdom)”, OECD Education Working Papers 2021. 10.1787/d99e8c0a-en
- “The Learning Gain Over One School Year Among 15-Year-Olds: An International Comparison Based on PISA”, *Labour Economics* 84 2023. 10.1016/j.labeco.2023.102365
- Baird, Sarah, Craig McIntosh, and Berk Özler, “Cash or Condition? Evidence from a Cash Transfer Experiment”, *The Quarterly Journal of Economics* 126 2011, 1709–1753. 10.1093/qje/qjr032
- Bambrick-Santoyo, Paul, *A Principal Manager’s Guide to Leverage Leadership 2.0: How to Build Exceptional Schools Across Your District*, ed. (John Wiley & Sons, Ltd, 2018). 10.1002/9781119548508.fmatter
- Banerjee, Abhijit V. et al., “Remedying Education: Evidence from Two Randomized Experiments in India”, *The Quarterly Journal of Economics* 122 2007, 1235–1264. 10.1162/qjec.122.3.1235
- Barrera-Osorio, Felipe and Leigh L. Linden, *The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia*, 2009. 10.1596/1813-9450-4836
- Barro, Robert J, “Human Capital and Growth”, *American Economic Review* 91 2001, 12–17. 10.1257/aer.91.2.12
- Barros, Ricardo Paes et al., “Promovendo o Desempenho Educacional Via Melhorias na Gestão Escolar: O Caso do Programa Jovem de Futuro”, *Pesquisa e Planejamento Econômico* 51 2021, 9–44. https://repositorio.ipea.gov.br/bitstream/11058/11668/1/PPE_v51_n03_Artigo1_promovendo_o_desempenho.pdf
- Barros, Ricardo Paes de et al., *Assessment of the Impact of the Jovem de Futuro Program on Learning*, World Bank, 2019. <https://documents1.worldbank.org/curated/en/825101561723584640/pdf/Assessment-of-the-Impact-of-the-Jovem-de-Futuro-Program-on-Learning.pdf>
- Becker, Gary S, Kevin M Murphy, and Robert Tamura, “Human capital, fertility, and economic growth”, *Journal of Political Economy* 98 1990, S12–S37.
- Beg, Sabrin A, Anne E Fitzpatrick, and Adrienne Lucas, *Managing to Learn*, Working Paper 31757, National Bureau of Economic Research, 2023. 10.3386/w31757
- Bettinger, Eric, “Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores”, *The Review of Economics and Statistics* 94 2012, 686–698. 10.1162/REST_a_00217
- Blimpo, Moussa P. and David K. Evans, *School-Based Management and Educational Outcomes: Lessons From a Randomized Field Experiment*, 2013.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, *Management as a technology?*, NBER Working Paper 22327, National Bureau of Economic Research, 2016. 10.3386/w22327
- Bloom, Nicholas and John Van Reenen, “Measuring and Explaining Management Practices Across Firms and Countries”, *The Quarterly Journal of Economics* 122 2007, 1351–1408. 10.1162/qjec.2007.122.4.1351
- “New Approaches to Surveying Organizations”, *The American Economic Review* 100 2010, 105–109. 10.1257/aer.100.2.105
- Bloom, Nicholas et al., “Does Management Matter? Evidence from India”, *The Quarterly Journal of Economics* 128 Nov. 2012, 1–51. 10.1093/qje/qjs044

- Bloom, Nicholas et al., “Management Practices Across Firms and Countries”, *Academy of Management Perspectives* 26 2012, 12–33. 10.5465/amp.2011.0077
- Bloom, Nicholas et al., “JEEA-FBBVA Lecture 2013: The New Empirical Economics of Management”, *Journal of the European Economic Association* 12 2014, 835–876. 10.1111/jeea.12094
- Bloom, Nicholas et al., “Does Management Matter in Schools?”, *The Economic Journal* 125 2015, 647–674. 10.1111/eoj.12267
- Bloom, Nicholas et al., “What Drives Differences in Management Practices?”, *American Economic Review* 109 May 2019, 1648–83. 10.1257/aer.20170491
- Bloom, Nicholas et al., “Do Management Interventions Last? Evidence from India”, *American Economic Journal: Applied Economics* 12 Apr. 2020, 198–219. 10.1257/app.20180369
- Borkum, Evan, Fang He, and Leigh L. Linden, *The Effects of School Libraries on Language Skills: Evidence from a Randomized Controlled Trial in India*, NBER Working Paper 18183, National Bureau of Economic Research, 2012. 10.3386/w18183
- Borman, Geoffrey D, Robert E Slavin, and Alan C. K Cheung, “Final Reading Outcomes of the National Randomized Field Trial of Success For All”, *American Educational Research Journal* 44 2007, 701–731. 10.3102/0002831207306743
- Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, “The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico”, *Journal of Political Economy* 126 2018, 635–687. 10.1086/696154
- Burde, Dana and Leigh L. Linden, “Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools”, *American Economic Journal: Applied Economics* 5 2013, 27–40. 10.1257/app.5.3.27
- Card, David, “Chapter 30 - The Causal Effect of Education on Earnings”, in *Handbook of Labor Economics*, vol. 3 (Elsevier, 1999). 10.1016/S1573-4463(99)03011-4
- Chaisemartin, Clément de and Jaime Ramirez-Cuellar, “At What Level Should One Cluster Standard Errors in Paired and Small-Strata Experiments?”, *American Economic Journal: Applied Economics* 16 2024, 193–212. 10.1257/app.20210252
- Chetty, Raj, John N Friedman, and Jonah E Rockoff, “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood”, *The American Economic Review* 104 2014, 2633–2679. 10.1257/aer.104.9.2633
- Cristia, Julian et al., “Technology and Child Development: Evidence from the One Laptop per Child Program”, *American Economic Journal: Applied Economics* 9 2017, 295–320. 10.1257/app.20150385
- Curto, Vilsa E. and Roland G. Fryer, “The Potential of Urban Boarding Schools for the Poor: Evidence from SEED”, *Journal of Labor Economics* 32 2014, 65–93. 10.1086/671798
- Davies, Neil M et al., “The Causal Effects of Education on Health Outcomes in the Uk Biobank”, *Nature Human Behaviour* 2 2018, 117–125. 10.1038/s41562-017-0279-y
- Dhaliwal, Iqbal et al., “Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education”, in *Education Policy in Developing Countries* (University of Chicago Press, 2013). 10.7208/9780226078854-008
- Dobbie, Will and Roland G Fryer, “Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children’s Zone”, *American Economic Journal: Applied Economics* 3 2011, 158–187. 10.1257/app.3.3.158

- Dobbie, Will and Roland G. Fryer, “Getting Beneath the Veil of Effective Schools: Evidence From New York City”, *American Economic Journal: Applied Economics* 5 2013, 28–60. 10.1257/app.5.4.28
- Doran, George T et al., “There’s a Smart Way to Write Management’s Goals and Objectives”, *Management Review* 70 1981, 35–36.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer, “Peer Effects, Teacher Incentives, and The Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”, *The American Economic Review* 101 2011, 1739–1774. 10.1257/aer.101.5.1739
- “School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools”, *Journal of Public Economics* 123 2015, 92–110. 10.1016/j.jpubeco.2014.11.008
- Duflo, Esther, Rema Hanna, and Stephen Ryan, “Incentives Work: Getting Teachers to Come to School”, *The American Economic Review* 102 2012, 1241–1278. 10.1257/aer.102.4.1241
- Feinstein, Alvan R. and Domenic V. Cicchetti, “High Agreement but Low Kappa: I. The Problems Of Two Paradoxes”, *Journal of Clinical Epidemiology* 43 1990, 543–549. 10.1016/0895-4356(90)90158-L
- Fryer Roland G., Jr., *Management and Student Achievement: Evidence from a Randomized Field Experiment*, NBER Working Paper 23437, National Bureau of Economic Research, 2017. 10.3386/w23437
- “The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments”, in *Handbook of Field Experiments*, vol. 2, Banerjee, Abhijit V. and Esther Duflo, eds. (Amsterdam: North-Holland, 2017).
- Fryer, Roland G, “Financial Incentives and Student Achievement: Evidence From Randomized Trials”, *The Quarterly Journal of Economics* 126 2011, 1755–1798. 10.1093/qje/qjr045
- “Injecting Charter School Best Practices Into Traditional Public Schools: Evidence From Field Experiments”, *The Quarterly Journal of Economics* 129 2014, 1355–1408. 10.1093/qje/qju011
- Gerber, Alan et al., “Reporting Guidelines for Experimental Research: A Report from the Experimental Research Section Standards Committee”, *Journal of Experimental Political Science* 1 2014, 81–98. 10.1017/xps.2014.11
- Gerber, Alan S, *Field Experiments: Design, Analysis, and Interpretation*, 2nd ed. (New York, N.Y.: W. W. Norton & Company, 2012).
- Gerber, Alan S. et al., “Reporting Balance Tables, Response Rates and Manipulation Checks in Experimental Research: A Reply from the Committee that Prepared the Reporting Guidelines”, *Journal of Experimental Political Science* 2 2015, 216–229. doi:10.1017/XPS.2015.20
- Gibbons, Robert and Rebecca Henderson, “Relational Contracts and Organizational Capabilities”, *Organization Science* 23 2012, 1350–1364. 10.1287/orsc.1110.0715
- Glazerman, Steven, Daniel Mayer, and Paul Decker, “Alternative Routes to Teaching: The Impacts of Teach for America on Student Achievement and Other Outcomes”, *Journal of Policy Analysis and Management* 25 2006, 75–96. 10.1002/pam.20157
- Glazerman, Steven et al., *Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment*, NCEE Report 2014-4003, National Center for Education Evaluation and Regional Assistance, 2013. 10.1037/e537912013-001
- Glewwe, Paul, Nauman Ilias, and Michael Kremer, “Teacher Incentives”, *American Economic Journal: Applied Economics* 2 2010, 205–227. 10.1257/app.2.3.205

- Glewwe, Paul, Michael Kremer, and Sylvie Moulin, “Many Children Left Behind? Textbooks and Test Scores in Kenya”, *American Economic Journal: Applied Economics* 1 2009, 112–35. 10.1257/app.1.1.112
- Gosnell, Greer K., John A. List, and Robert D. Metcalfe, “The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains”, *The Journal of Political Economy* 128 2020, 1195–1233.
- Gwet, K.L., *Handbook of Inter-Rater Reliability*, 5th Edition: The Definitive Guide to Measuring The Extent of Agreement Among Raters (Advanced Analytics, LLC, 2021).
- Hayes, Richard J. and Lawrence H. Moulton, *Cluster Randomised Trials* (CRC press, 2017).
- Hoyos, Rafael de, Alejandro J Ganimian, and Peter A Holland, “Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity Building for Public Schools in Argentina”, *The World Bank Economic Review* 35 Nov. 2019, 499–520. 10.1093/wber/lhz026
- Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge University Press, 2015). 10.1017/CBO9781139025751
- Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP), Educational financial indicators, INEP, 2021. <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-financeiros-educacionais>
- J-PAL, Conducting cost-effectiveness analysis (CEA), J-PAL, 2020. <https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea>
- Kahan, Brennan C et al., “Estimands in Cluster-Randomized Trials: Choosing Analyses that Answer the Right Question”, *International Journal of Epidemiology* 52 2023, 107–118. 10.1093/ije/dyac131
- Kayembe, Mutamba T. et al., “Imputation of Missing Covariates in Randomized Controlled Trials with Continuous Outcomes: Simple, Unbiased and Efficient Methods”, *Journal of Biopharmaceutical Statistics* 32 2022, 717–739. 10.1080/10543406.2021.2011898
- Kraft, Matthew A., “Interpreting Effect Sizes of Education Interventions”, *Educational Researcher* 49 2020, 241–253. 10.3102/0013189X20912798
- Kremer, Michael, Edward Miguel, and Rebecca Thornton, “Incentives to Learn”, *The Review of Economics and Statistics* 91 2009, 437–456. 10.1162/rest.91.3.437
- Krueger, Alan B., “Experimental Estimates of Education Production Functions”, *The Quarterly Journal of Economics* 114 1999, 497–532. 10.1162/003355399556052
- “Economic Considerations and Class Size”, *The Economic Journal* (London) 113 2003, F34–F63. 10.1111/1468-0297.00098
- Landis, J. Richard and Gary G. Koch, “The Measurement of Observer Agreement for Categorical Data”, *Biometrics* 33 1977, 159–174. 10.2307/2529310
- Lassibille, Gérard et al., “Managing for Results in Primary Education in Madagascar: Evaluating the Impact of Selected Workflow Interventions”, *The World Bank Economic Review* 24 2010, 303–329. 10.1093/wber/lhq009
- Lin, Winston, “Agnostic Notes on Regression Adjustments to Experimental Data: Re-examining Freedman’s Critique”, *The Annals of Applied Statistics* 7 2013, 295–318. 10.1214/12-AOAS583
- Lucas, Robert E., “On the Mechanics of Economic Development”, *Journal of Monetary Economics* 22 1988, 3–42. 10.1016/0304-3932(88)90168-7

- Morrow-Howell, Nancy et al., *Evaluation of Experience Corps: Student Reading Outcomes*, 2009.
- Muralidharan, Karthik and Paul Niehaus, “Experimentation at Scale”, *The Journal of Economic Perspectives* 31 2017, 103–124. 10.1257/jep.31.4.103
- Muralidharan, Karthik and Abhijeet Singh, *Improving Public Sector Management at Scale? Experimental Evidence on School Governance India*, NBER Working Paper 28129, National Bureau of Economic Research, 2020. 10.3386/w28129
- Muralidharan, Karthik and Venkatesh Sundararaman, “The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India”, *The Economic Journal* 120 2010, F187–F203. 10.1111/j.1468-0297.2010.02373.x
- Mutz, Diana C., Robin Pemantle, and Philip Pham, “The Perils of Balance Testing in Experimental Design: Messy Analyses of Clean Data”, *The American Statistician* 73 2019, 32–42. 10.1080/00031305.2017.1322143
- Neal, Derek A. and William R. Johnson, “The Role of Premarket Factors in Black-White Wage Differences”, *The Journal of Political Economy* 104 1996, 869–895. 10.1086/262045
- Negi, Akanksha and Jeffrey M. Wooldridge, “Revisiting Regression Adjustment in Experiments with Heterogeneous Treatment Effects”, *Econometric Reviews* 40 2021, 504–534. 10.1080/07474938.2020.1824732
- Nguyen, Trang, *Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar*, Massachusetts Institute of Technology, 2008. <https://www.povertyactionlab.org/sites/default/files/documents/Nguyen%5C%202008.pdf>
- OECD, *Education at a Glance 2021* (Paris: OECD Publishing, 2021). 10.1787/b35a14e5-en
- *PISA 2022 Results (Volume I)* (Paris: OECD Publishing, 2023). 10.1787/53f23881-en
- Oreopoulos, Philip and Kjell G Salvanes, “Priceless: The Nonpecuniary Benefits of Schooling”, *The Journal of Economic Perspectives* 25 2011, 159–184. 10.1257/jep.25.1.159
- Pradhan, Menno et al., “Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia”, *American Economic Journal. Applied Economics* 6 2014, 105–126. 10.1257/app.6.2.105
- Pritchett, Lant, *The Rebirth of Education: Schooling Ain’t Learning* (Brookings Institution Press, 2013). <http://www.jstor.org/stable/10.7864/j.ctt1gpcdb8>
- Project Management Institute, *Agile Practice Guide* (Newtown Square, PA: Project Management Institute, Inc. (PMI), 2017).
- *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition and The Standard for Project Management* (Chicago: Project Management Institute, 2021).
- Puma, Michael et al., “Head Start Impact Study. Final Report.”, Administration for Children & Families 2010.
- Romero, Mauricio et al., “Direct vs Indirect Management Training: Experimental Evidence from Schools in Mexico”, *Journal of Development Economics* 154 2022, 102779. 10.1016/j.jdeveco.2021.102779
- Senn, Stephen, “Testing for Baseline Balance in Clinical trials”, *Statistics in Medicine* 13 1994, 1715–1726. 10.1002/sim.4780131703
- Sloczynski, Tymon, “Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights”, *The Review of Economics and Statistics* 104 2022, 501–509. 10.1162/rest_a.00953

- Somers, Marie-Andrée et al., The Enhanced Reading Opportunities Study Final Report: The Impact of Supplemental Literacy Courses for Struggling Ninth-Grade Readers, NCEE Report 2010-4021, National Center for Education Evaluation and Regional Assistance, 2010. <https://ies.ed.gov/ncee/pubs/20104021/pdf/20104021.pdf>
- Tavares, Priscilla Albuquerque, “The Impact of School Management Practices on Educational Performance: Evidence from Public Schools in São Paulo”, *Economics of Education Review* 48 2015, 1–15. 10.1016/j.econedurev.2015.05.002
- TCMRio, Relatório e Projeto de Parecer Prévio: Contas de Governo 2019, Tribunal de Contas do Município do Rio de Janeiro, 2020. https://www.tcmrio.tc.br/Noticias/14466/Peca140_RelatorioProjetoParecerPrevio_LAG.pdf
- Thaler, Richard H. and Cass R. Sunstein, *Nudge*, final edition (London: Allen Lane, 2021).
- World Bank, *World Development Report 2018: Learning to Realize Education’s Promise*, 2018. 10.1596/978-1-4648-1096-1
- Zhao, Anqi and Peng Ding, “To Adjust or not to Adjust? Estimating the Average Treatment Effect in Randomized Experiments with Missing Covariates”, *Journal of the American Statistical Association* 0 2022, 1–11. 10.1080/01621459.2022.2123814