# DISCUSSION PAPER SERIES

DP14069
(v. 2)

## UNDERSTANDING SCHOOL MANAGEMENT WITH PUBLIC DATA: A NEW MEASUREMENT APPROACH AND APPLICATIONS

Clare Leaver, Renata Lemos and Daniela Scur

**LABOUR ECONOMICS, MACROECONOMICS AND GROWTH AND PUBLIC ECONOMICS**

CEPR

ISSN 0265-8003

# UNDERSTANDING SCHOOL MANAGEMENT WITH PUBLIC DATA: A NEW MEASUREMENT APPROACH AND APPLICATIONS

*Clare Leaver, Renata Lemos and Daniela Scur*

Discussion Paper DP14069
First Published 21 October 2019
This Revision 28 July 2024

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

# UNDERSTANDING SCHOOL MANAGEMENT WITH PUBLIC DATA: A NEW MEASUREMENT APPROACH AND APPLICATIONS


## Abstract

Why do students learn more in some schools than others? One consideration receiving growing attention is school management. To study this, researchers need to be able to measure school management accurately and cheaply at scale. We introduce a new approach to measuring management practices using existing public data and exemplify the methodology with OECD's PISA and Brazil's Prova Brasil. Both indices show a strong, positive relationship between management and learning. But why do management practices impact student learning? We propose a new theoretical framework that maps school management to school functioning via intermediate outcomes relating to teachers and households, and from this to student learning. An application of our methods highlights that both principals and practices matter for student outcomes.



JEL Classification: M5, I2, J3

Keywords: Management, Teacher selection, Teacher incentives, Cross-country



Clare Leaver - clare.leaver@bsg.ox.ac.uk
*Blavatnik School of Government, University of Oxford and CEPR*

Renata Lemos - rlemos@worldbank.org
*World Bank and CEP-LSE and CEPR*

Daniela Scur - daniela.scur@gmail.com
*Cornell University, Dyson School of Applied Economics and Management, CEP-LSE and CEPR*



Acknowledgements
We thank Melissa Adelman, Chris Barrett, Nick Bloom, Gustavo Bobonis, Alex Eble, Vicente Garcia, John Hoddinott, Abhijeet Singh and Nikki Shure for helpful discussions and comments. We also thank participants and discussants at the RISE Annual Conference 2018, SIOE 2019, Boston University, Blavatnik School of Government, Cornell University, and World Bank's Regional Study authors' workshop. We thank Jose Mola, Raissa Ebner, Maria Jose Vargas, Claudia Rivas, and Ildo Lautharte for their help and support throughout this project, and we are especially thankful to Fabiano Dal-Ri and Rafael Rocha-Lima Bentes for excellent research assistance. Leaver is grateful for the hospitality of Toulouse School of Economics, 2018-19.

# Understanding school management with public data: A new measurement approach and applications

Clare Leaver* University of Oxford

Renata Lemos† World Bank

Daniela Scur‡ Cornell University

**This draft:**
July 28, 2024

**Abstract**

Why do students learn more in some schools than others? One consideration receiving growing attention is school management. To study this, researchers need to be able to measure school management accurately and cheaply at scale. We introduce a new approach to measuring management practices using existing public data and exemplify the methodology with OECD's *PISA* and Brazil's *Prova Brasil*. Both indices show a strong, positive relationship between management and learning. But *why* do management practices impact student learning? We propose a new theoretical framework that maps school management to school functioning via intermediate outcomes relating to teachers and households, and from this to student learning. An application of our methods highlights that *both* principals and practices matter for student outcomes.

**Keywords:** management, teacher selection, teacher incentives, cross-country
**JEL codes:** M5, I2, J3

---

* **Acknowledgements:** We thank Melissa Adelman, Chris Barrett, Nick Bloom, Gustavo Bobonis, Alex Eble, Vicente Garcia, John Hoddinott, Abhijeet Singh and Nikki Shure for helpful discussions and comments. We also thank participants and discussants at the RISE Annual Conference 2018, SIOE 2019, Boston University, Blavatnik School of Government, Cornell University, and World Bank's Regional Study authors' workshop. We thank José Mola, Raissa Ebner, Maria José Vargas, Claudia Rivas, and Ildo Lautharte for their help and support throughout this project, and we are especially thankful to Fabiano Dal-Ri and Rafael Rocha-Lima Bentes for excellent research assistance. Leaver is grateful for the hospitality of Toulouse School of Economics, 2018-19.

*CEPR and Toulouse School of Economics. E-mail: clare.leaver@bsg.ox.ac.uk.

†CEPR; CEP-LSE; IGC. E-mail: rlemos@worldbank.org.

‡Corresponding Author. CEPR & CEP-LSE. E-mail: dscur@cornell.edu. Dyson School of Applied Economics and Management, 137 Reservoir Ave, Ithaca, NY 14853, United States.

# 1 Introduction

Despite global calls for improvements in education, progress towards learning for all is slow. This deficit is particularly pronounced for poor children and children in low-income countries (Akmal and Pritchett, 2019; Cullen et al., 2013). But why do some students learn more in some schools than others? While there are many contributing factors at system, school, and household-level, one consideration receiving growing attention is school management practices. These practices are distinct from principal characteristics and leadership, and refer to the processes and practices used by principals day-to-day as they run their schools (World Bank, 2018). However, researchers and practitioners interested in this issue face two key challenges in accounting for the role of management practices in their work: (a) understanding *how* school management leads to better student learning; and (b) how to *measure* school management accurately and cost-effectively at scale and across contexts.

In this paper we address both challenges. We first develop a theoretical framework that maps school management to school functioning (intermediate outcomes relating to teachers and households) and from this into student learning. We outline theoretical causal pathways to illuminate the black box this topic currently sits in, and demonstrate that the predictions of the model are supported by international data. This provides validation for a framework that can be (and indeed has been) used by researchers with access to the data necessary to explore the causal impact of management practices on school outcomes. Prem and Munoz (2023), for example, build on our framework to show "*how* principals matter" using administrative data and a natural experiment on principal recruitment in Chile.

In parallel, we develop a new approach to measurement that can, in principle, be used with any existing public dataset containing information about school management. We illustrate the methodology with two example datasets: the OECD's Programme for International Student Assessment (PISA) and the Brazilian Prova Brasil assessments and surveys. We show how questions from these public surveys can be classified into management practices and coded similarly to the rubric from well-established management practices surveys (c.f. Bloom et al., 2015; Buffington et al., 2017) and built into a school management index. Our PISA-based index covers over 15,000 schools across 65 countries, and our Prova Brasil-based index covers nearly all public schools in Brazil (over 72,000). We supplement these management indices by using questions relating to teacher shortages, motivation and effort, and household engagement, to construct measures of school functioning, both for PISA and Prova Brasil. We validate these new indices and show that they are correlated with student learning, conditional on a rich set of controls.

We hope there will be myriad uses of these new indices and the methodology for building them. We demonstrate the potential value of this new data with applications focusing on unpacking the importance of practices and principals for student performance. First, we build on the results from Akhtari et al. (2022) that document "upheaval" in school leadership due to changes in municipal governments following close elections lead to a subsequent decline in student performance.[1] We show that this decline is likely working – at least partially – through a management practices channel. Further, we show that management practices and principal characteristics contribute similarly to attenuating the negative partial effect of principal turnover on student learning; i.e., people matter, but so do structured practices. Second, we explore *how* simple structures can have such an important impact on student learning, testing the predictions of our model and finding empirical support for the pathways we propose.

Our new measurement approach contributes to the literature on the role of managers and management practices in determining establishment performance. Results for firms are consistent: managers and management practices matter for productivity (e.g. Bloom et al., 2019; Bloom and Van Reenen, 2007; Giorcelli, 2019; Scur et al., 2021; Syverson, 2011) and labor flows (Bender et al., 2018; Cornwell et al., 2021). This relationship has also been documented in the public sector, including schools in both high-state-capacity contexts (Bloom et al., 2015; Fryer, 2014, 2017) and low-state-capacity contexts (Crawfurd, 2017; Lemos et al., 2021; Romero et al., 2020).[2] To date, however, the number of schools and countries studied has been relatively small, primarily as a result of data limitations (Adelman and Lemos, 2021). Our approach enables researchers to work with substantially larger datasets and, as we illustrate in our first application, facilitates quasi-experimental methods for evaluating the impact of turnover in political and organizational leadership. A further benefit is that researchers using experimental methods now have a larger set of "benchmarkable" questions on management practices to consider for inclusion in baseline and endline surveys (as in Crawfurd et al. (2024)).

A related literature considers the role of education systems and institutions in determining student performance across countries (Wössmann, 2016). PISA is a commonly used

[1] Leadership change is notoriously disruptive, though the wider literature on the disruption costs of leadership change often focuses on CEOs (Bandiera et al., 2020; Bertrand and Schoar, 2003; Huber et al., 2021a). Causal evidence in the CEO literature is scarce, typically relying on rare events such as CEO deaths or wars. Our analysis exploits a commonplace event, political turnover, and shows that leadership change in schools leads to deterioration in both management practices and establishment performance.

[2] In addition to schools, sectors studied include: universities (McCormack et al., 2014), healthcare facilities (Bloom et al., 2017, 2015), social programs (Delfgaauw et al., 2011; McConnell et al., 2009), the civil service (Fenizia, 2022; Rasul and Rogger, 2016) and state-owned enterprises (Kala, 2019).

dataset and researchers have looked at this issue through the lens of autonomy (Hanushek et al., 2013; Wössmann et al., 2007), competition (West and Wössmann, 2010), student tracking (Hanushek and Wössmann, 2006; Ruhose and Schwerdt, 2016), external exams (Wössmann, 2005), and instructional time (Lavy, 2015). Our new indices, especially the PISA-based ones, enable researchers to consider school management in such studies and across a larger number of countries.

Our theoretical framework also makes a modest contribution to the literature on incentives and selection in public sector organizations (see Dal Bó and Finan (2020) and Finan et al. (2017) for recent reviews). Lazear (2003), Dohmen and Falk (2010) and Leaver et al. (2021) emphasise the potential selection margin of teacher performance pay. A selection margin also features in the dynamic occupational model of Rothstein (2015) and the Roy model of Biasi (2021). Our contribution in this paper is to focus on other aspects of school management (rather than performance pay) and to provide an intuitive decomposition of the impact of these practices on student learning.

## 2 Why might management matter in schools?

It is now well established that good management practices in schools are associated with better student learning outcomes, but *why* these practices matter and the channels they are working through remain a black box. We develop a simple theoretical framework in which good management practices drive student learning by improving school functioning, deriving testable predictions that can then be taken to the data.

The framework focuses on teachers. Our aim is not to provide a theoretical contribution *per se*, but rather to formalize intuitions around teacher incentive and selection mechanisms and their relationship to management practices and student performance.[3] We take wider system-level factors — in particular hiring and firing autonomy, admissions autonomy and competition between schools — as given and assume that teachers and students make choices within the confines of this environment.

Real-world education systems are diverse, particularly in terms of the type of private sector offerings. In some contexts, private schools target affluent households, and jobs in private schools are seen as more attractive than jobs in public schools, typically providing some form of performance-based compensation. In other contexts, there has been a growth of 'low-cost' private schools that deliberately cater for the lower end of the income distribution and, in these settings, jobs in the public sector often confer significant rents relative to the

[3] In this sense, we build on the model sketched in the appendix to Leaver et al. (2021).

private sector. In view of this diversity, we restrict our analysis to the sector that attends to the largest share of students across countries, namely public schools.

## 2.1 Theoretical framework

We focus on a teacher who must decide whether to accept a job offer in her assigned public school, or decline it and apply to a private school or the outside sector. The teacher is risk neutral and cares about her compensation $w$ and effort $e$. When working in the education sector, the teacher's preferences are $w - (e^2 - c\, e)$. The parameter $c$ captures her *intrinsic motivation*. This is because for $e < c/2$ she derives a marginal benefit from exerting an extra unit of effort in teaching; it is only when $e > c/2$ that effort costs kick in. We assume that $c = \tau + \Delta$. The first component $\tau$ denotes the teacher's baseline intrinsic motivation. This can be thought of as the realization of a random variable. The teacher observes this realization perfectly, while (at the time of hiring) employers observe nothing. The second component $\Delta$ is a motivational increment that is determined by the management practices in the teacher's chosen school. When working in the other sector, the teacher's preferences are simply $w - e^2$; intrinsic motivation plays no role. We abstract from student heterogeneity and focus on a representative household (student plus parents). This household cares only about its effort level $a$, and has preferences $-(a^2 - \gamma\, a)$. The parameter $\gamma$ is a motivational increment that is also determined by management practices.

Let $y_1$ denote a representative student's learning outcome in a school that hires the teacher, and $y_0$ denote a representative student's learning outcome in a school that does not hire the teacher. To the extent that teachers contribute to learning, one would expect $y_1 > y_0$. We capture this in a simple way by assuming $y_1 = \theta e + a + \varepsilon$ and $y_0 = a + \varepsilon$. If the teacher is not hired by a school but instead chooses to work in the outside sector, her performance is $z = \theta e + \varepsilon$. The component $\theta$ denotes the teacher's *ability*. This can also be thought of as the realization of a random variable, assumed to be independent of $\tau$. The teacher observes this realization perfectly, while (at the time of hiring) employers observe nothing. Draws of the error term $\varepsilon$ are independent across employments. We assume throughout that $\varepsilon$ is mean zero and distributed $U\left[\underline{\varepsilon}, \bar{\varepsilon}\right]$.

Public schools offer a wage of $G$. Private schools offer a base wage of $W$ plus a bonus $B$ if the teacher's performance exceeds a threshold $\bar{y}$. The outside sector offers a low base wage (normalized to zero) and a bonus $\beta$ if performance exceeds a threshold $\bar{z}$.

We assume that management has three effects. The first relates to teacher motivation: good management practices enable managers to cultivate the intrinsic motivation of their staff, increasing $\Delta$. The second relates to compensation: good management practices free up resources and enable managers to offer a higher level of pay (potentially in hedonic form). The third relates to household effort: good management practices help to create a stimulating environment for students and parents, increasing $\gamma$. Our interest lies in establishing how these three effects translate into student learning. We do not model the government's assignment rule, or the school principal's choice of management practices. For simplicity, we classify schools as either high or low management. In a high management school, $\Delta$, base pay, and $\gamma$ are all higher than in a low management school. Below is a summary of this description of the model.

1. Nature chooses the teacher's two-dimensional type. This realization $(\tau, \theta)$ is observed by the teacher but not by employers.

2. Employers announce management practices and compensation schemes.

3. The teacher is assigned (by government) to a public school and decides whether to accept this post or decline it and apply either to a private school or the outside sector.

4. Having made an occupational choice, the teacher chooses an effort level. Simultaneously, if the teacher is in the education sector, households choose effort levels.

5. A performance metric is realized. The teacher is rewarded in accordance with the compensation scheme announced at Stage 2.

## 2.2 Mechanisms

We use this framework to show how public schools with good management can produce better student outcomes. Specifically, we compare outcomes in a high management public school with outcomes in a low management public school, assuming both compete with a high management private school and the outside sector. The index $i = L, H$ denotes the quality of management in these public schools.

Public school $i$ hires its assigned teacher if, given her $(\theta, \tau)$ type, she expects to receive a higher payoff from teaching in this school compared to a high management private school or working in the outside sector. We use the notation $\mathcal{T}^i$ to denote the set of $(\theta, \tau)$ types that can be hired to this school. The expected learning outcome of a representative student (*ex ante*, prior to occupational and effort choices) is

$$\mathrm{E}\left[y^i\right] = \mathrm{E}\left[y_1^i \cdot 1_{\{(\theta,\tau)\in\mathcal{T}^i\}}\right] + \mathrm{E}\left[y_0^i \cdot 1_{\{[(\theta,\tau)\notin\mathcal{T}^i\}}\right],$$

where $1_{\{(\theta,\tau)\in\mathcal{T}^i\}}$ and $1_{\{[(\theta,\tau)\notin\mathcal{T}^i\}}$ are indicator functions for the hiring and not hiring events. In keeping with the empirical application, we will refer to $\mathrm{E}\left[y^i\right]$ as the expected test score in school $i$.

The difference in expected test scores between high and low management public schools can be written as

$$\mathrm{E}\left[y^H\right] - \mathrm{E}\left[y^L\right] = \mathrm{E}\left[y_1^H \cdot 1_{\{(\theta,\tau)\in\mathcal{T}^H\}}\right] - \mathrm{E}\left[y_1^L \cdot 1_{(\theta,\tau)\in\mathcal{T}^L\}}\right] + \mathrm{E}\left[y_0^H \cdot 1_{\{(\theta,\tau)\notin\mathcal{T}^H\}}\right] - \mathrm{E}\left[y_0^L \cdot 1_{\{(\theta,\tau,)\notin\mathcal{T}^L\}}\right].$$

In the Appendix, we derive teacher and household effort in high and low management public schools. These optimal choices are $e^i = \frac{\tau+\Delta^i}{2}$ and $a^i = \frac{\gamma^i}{2}$ for $i = L, H$. Substituting for these expressions, we can decompose the difference in expected test scores as

$$\mathrm{E}\left[y^H\right] - \mathrm{E}\left[y^L\right] = \underbrace{\mathrm{E}\left[\theta\left(\frac{\Delta^H - \Delta^L}{2}\right) \cdot 1_{\{(\theta,\tau)\in\mathcal{T}^H\}}\right]}_{\text{teacher incentives}} + \underbrace{\mathrm{E}\left[\theta\left(\frac{\tau + \Delta^L}{2}\right) \cdot \left(1_{\{(\theta,\tau)\in\mathcal{T}^H\}} - 1_{\{(\theta,\tau)\in\mathcal{T}^L\}}\right)\right]}_{\text{teacher selection}} + \underbrace{\frac{\gamma^H - \gamma^L}{2}}_{\text{household incentives}}. \quad (1)$$

The first term on the RHS of equation (1) is what we term the *teacher incentive effect* of good management. Here, we compare the expected teacher contribution to the test score outcome in a high management public school, in the event that the teacher is hired to such a school, against the expected teacher contribution in a low management public school, in the counterfactual event that the teacher is hired to a high management public school. In this way, we hold the set of $(\theta, \tau)$ types fixed and just consider how the incentive environment for the teacher contributes to test scores. This expression shows that the test score is higher in a high management public school, in part, because good management practices increase the intrinsic motivation of any given $(\theta, \tau)$ type of teacher, who then exerts more effort than she would in a low management public school.

The second term in equation (1) captures what we term the *teacher selection effect* of good management practices. Here, we compare the expected teacher contribution to the test score outcome in a low management public school, in the event that the teacher is hired to such a school, against the expected teacher contribution in a low management public school, in the counterfactual event that the teacher is hired to a high management school. The test score is higher in a high management public school, in part, because good management practices encourage better $(\theta, \tau)$ types to select in, and these types exert more effort and are of greater ability than would be the case in a low management school.

Figure 1 provides an illustration of this teacher selection effect.[4] The unshaded area in the top panel depicts the set of $(\theta, \tau)$ types that are hired by a high management public school, while the unshaded area in the bottom panel depicts the set of $(\theta, \tau)$ types that are hired by a low management public school. Note that the high management public school hires *more* types than the low management public school: the unshaded area is larger in the top panel relative to the bottom panel. It also hires *better* types: average $\theta$ and average $\tau$, shown by the $(x, y)$ coordinates of the blue dot, are higher relative to the bottom panel.[5] The third term in equation (1) captures what we term the *household incentive effect* of good management practices. We see from this expression that the test score is higher in a high management public school, in part, because good management practices increase the motivation of parents and students who then exert more effort than they would in a low management public school.

In summary, our theoretical framework proposes three mechanisms — teacher incentives, teacher selection and household incentives — that could explain the positive correlation between management scores and student learning outcomes apparent in much of the earlier work on this topic. If these mechanisms are correct, then we should see behavioural responses in school functioning. Below, we set out the predictions from our model, all of which are testable implications that can be explored using detailed datasets.

**Prediction 1: Teacher shortages.** The theoretical framework predicts that the probability of hiring the teacher in a high management public school is higher than the probability of hiring the teacher in a low management public school (via teacher selection).

**Prediction 2: Teacher motivation.** The theoretical framework predicts that the expected intrinsic motivation of a teacher hired to a high management public school is higher than the expected intrinsic motivation of a teacher hired to a low management public school (via teacher selection), at least in settings without low-cost private schools.

[4] Details of the construction of Figure 1 are provided in the Appendix. Bonus pay in the private education sector and in the outside sector is assumed to be higher than the public sector wage. In the Appendix, we also consider the case where public sector pay exceeds pay in the private education sector (so called 'low-cost' private schools).

[5] As we discuss in the Appendix, the prediction that the high management school hires better $\theta$ types is sensitive to parameter assumptions and, for instance, does not hold in our numerical example with 'low-cost' private schools.

**Prediction 3: Teacher effort.** The theoretical framework predicts that the expected effort level of a teacher hired to a high management public school is higher than the expected effort level of a teacher hired to a low management public school (via teacher selection and incentives).

**Prediction 4: Household effort.** The theoretical framework predicts that expected household effort in a high management public school is higher than expected household effort in a low management public school (via household incentives).

We use this theoretical framework to ground the development of indices of intermediate outcomes with public data. We will show that all these predictions have support in the two datasets that we build, suggesting that the causal pathways from the quality of management practices to student learning posited in the theory — selection and incentives within and beyond the school — are empirically plausible across a wide range of countries. While the evidence presented here is only correlational, it does however provide validation for the framework, which as Prem and Munoz (2023) have shown, can be used to explore the *causal* impact of management practices on school outcomes in other settings.

# 3 How to measure management in schools?

Until the early 2000s, management was typically viewed as an unmeasurable productivity shifter, relegated to the residual in performance regressions (Scur et al., 2021). In education research, elements of management were imbued in a black box of school or principal "fixed effects", often bundling manager and management factors. Böhlmark et al. (2016), for example, explore principal fixed effects, but even with their detailed Scandinavian datasets the authors note that "it is difficult to determine which are the principal characteristics that form the basis for successful school management."

Over the last two decades, improvements in survey methodology and data access have allowed for substantial advances in measurement of practices. In particular, the World Management Survey is a large, cross-country project that measures the adoption of management best practices in multiple sectors using a detailed and extensive, but also expensive, survey methodology. It was originally developed to measure the adoption of structured management practices in manufacturing firms (Bloom and Van Reenen, 2007) and following its success in this original sector the questionnaire and methodology were subsequently adapted to other sectors of the economy, including public sector organizations (e.g. Bloom et al., 2015, 2019; Lemos et al., 2024; McCormack et al., 2014). The rigorous data collection is based on double-blind, semi-structured interviews conducted by highly-trained analysts and monitored by supervisors experienced on the survey methodology.

The schools survey covers 20 topics, including: whether the school has standardization of instructional processes across classrooms while allowing for within-classroom personalization of learning; whether and how the school uses assessments and data; whether and how the school sets and uses targets and keeps track of progress and how principals handle good and bad teacher performance. For each topic, there is a scoring grid ranging from 1 (little to no structured management) to 5 (best practice), which serves as a guide to evaluate answers to questions during the interviews. The overall management index, which measures the level of adoption of structured management best practices, is simply the average of the scores for these 20 topics. The practices measured by the survey seem to matter: Bloom et al. (2015) show that their school management score is strongly positively correlated with school-level student outcomes across 6 countries (Brazil, Canada, India, Sweden, UK and US). They find a strong positive correlation for these countries: moving from the bottom to the top quartile of management is associated with a large increase in student learning outcomes, equivalent to approximately 0.4 standard deviations.

Core data for management in schools has been collected for 8 countries, for a total of over 1,800 schools. One of the key advantages of this method is that scoring of the practices adopted is done by an independent "analyst" who interviews dozens of principals (including across countries and languages, to ensure cross-country consistency). This allows for cross-country comparability, but it comes at a high cost (approximately USD400 per interview) as each data point is substantially labor intensive and it can be slow (it takes about 4 months to conduct a single country wave). In view of these costs, it is not well suited for every context and thus the inherent lack of scalability is one of the major weaknesses of the method (Scur et al., 2021).

Partially in response to this lack of scalability, a sister project was conceived at the US Census Bureau where the agency took a set of key practices that could be translated into a self-respondent questionnaire and created the Management and Organizational Practices Survey sent to over 35,000 manufacturing firms in the US (MOPS, Buffington et al. (2017)). The questionnaire was later translated and applied to 14 additional countries, demonstrating that it was possible to collect useful management practices data at scale. The main weaknesses of this self-respondent approach are that scale becomes necessary to account for the additional noise, and it is not possible to objectively account for cultural biases in responses.

We take the spirit of the MOPS exercise and explore existing datasets that include questions with information on management practices to build management indices. While there are limitations, this method has the advantage of not requiring additional resources: the dataset exists, is public, and often is continually updated by a large agency or group. This allows researchers to build and use management indices without large grants or projects, contributing to accessibility and reproducibility efforts. We outline below the process we follow to classify questions as a "management practice", and how we then use the subset of questions that have topics overlapping with the World Management Survey to benchmark the exercise. To illustrate the methodology, we use two popular datasets in education research: OECD's *PISA* and Brazil's *Prova Brasil*.

## 3.1 A decision tree for identifying management practices

We primarily use principal questionnaires, but include information from other questionnaires (teachers, parents, student) if they relate to the practice in question.[6] We review each survey question in the applicable questionnaire and ask four questions in succession, as in a decision tree. We use the following question from PISA 2012 to illustrate:

> Example 1: *"In your school, are assessments of students in <national modal grade for 15-year-olds> used for any of the following purposes?"*
>
> (a) *To inform parents about their child's progress*
>
> (b) *To make decisions about students' retention or promotion*
>
> (c) *To group students for instructional purposes*
>
> (d) *To compare the school to district, state, or national performance*
>
> (e) *To monitor the school's progress from year to year*
>
> (f) *To make judgments about teachers' effectiveness*
>
> (g) *To identify aspects of instruction or the curriculum that could be improved*
>
> (h) *To compare the school with other schools*
>
> – *PISA 2012, Question 16 from the Principal Questionnaire*

**Node 1: Is the survey question objective?** The first decision point is whether the survey question, as worded, can be objective or whether it is subjective. Objective questions ask about specific attributes, policies or processes. Subjective questions ask about sentiment, attitudes and opinions (without a standardized objective benchmark or "correct" answer).

[6] Full details to enable replication with these data sources are provided in the replication package.

As subjective information is hard to compare even within country or culture, we focus on objective information for management practices. In Example 1, the question and *all* the sub-options are objective – the question asks how student testing is used and for what purposes. It is *not* asking the principal's personal opinion on the use of these assessments, and the options are specific. An example of a question that would fail this first decision point is the first sub-option of PISA 2012, Question 30:

> Example 2: *"Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during <the last academic year>."*
>
> (a) *I pay attention to disruptive behaviour in classrooms.*
>
> – *PISA 2012, Question 30(a) from the Principal Questionnaire*

In the case of Example 2, the wording of option (a) is subjective because it relies on the respondent's understanding of "paying attention". Some respondents might consider simply waiting to hear about disruptive behavior as paying attention, while others might consider daily rounds of checking into classrooms as paying attention. As such, this question cannot provide *objective* information about this particular process at a school and thus we do not include it in our index.

**Node 2: Is the survey question asking about a process?** The second decision point is whether the information can help identify a process: a process is defined as referring to a series of actions. Specifically, we look for processes related to implementation (or, adoption), usage or monitoring of a practice. Attributes and policies, for example, are not processes. Attributes refer to characteristics of the organization (such as school size or ownership), while policies refer to written rules known to those concerned with the policy but do not offer information regarding its adoption. In Example 1, the main question and all sub-options are describing processes. An example of a question that would fail this second decision point is the first sub-option of PISA 2012 Question 32:

> Example 3: *"Which of the following measures aimed at quality assurance and improvement do you have in your school?"*
>
> (a) *Written specification of the school's curricular profile and educational goals.*
>
> – *PISA 2012, Question 32(a) from the Principal Questionnaire*

In the case of Example 3, the question is objective, but it is not describing a process. That is, it notes that there is a written specification but it does not describe any implementation or adoption processes of this directive. As such, it is providing information on a policy, not a practice, and thus we do not include it in our index.

**Node 3: Is the survey question linked to a management practice?** The third decision point is whether the objective description of a process is part of a *management practice* or not. We define a management practice as a set of processes employed by managers to lead and manage operations, people and/or resources in their organization. For example, pedagogical practices that are not linked directly with the organization of operations and people are not management practices. In the case of Example 1, the question and all the sub-options are objective processes linked to how principals use assessments to manage information flows (for example, sub-option a), manage students (sub-options b and c), monitor school performance and target-setting (sub-options d, e, g), monitor staff (sub-option f) and engage in continuous improvement (sub-option h). In PISA 2012, all questions we identified as being objective and describing a process were linked to a management practice. In principle, questions about, say, processes for school infrastructure would not have been included in a management index.

**Node 4: Does the topic addressed in the survey question have a WMS equivalent?** The final decision point *for our preferred index* is whether the topic addressed by the question has a WMS equivalent. This is because, as this is a new methodology, it is important to have a strong benchmark to ground the exploration. In the case of Example 1, all sub-options *except for sub-option (c)* pass this decision point. Sub-option (c) describes using student assessments for grouping students (i.e. "tracking"), which the WMS does not measure or ask about. To be clear, absent the goal of benchmarking, including such questions in a management index could be well worth doing. In the case of PISA 2012 there are only six questions that passed the first three decision points but failed this fourth one, and when we include them in the index the patterns remain similar.[7]

**Exploiting non-management questions.** One of the main advantages of large public datasets is that, beyond useful questions measuring management constructs, these datasets often also include a wealth of additional data on students, teachers, parents and the school. Beyond standard characteristics variables (such as school size, location or number of teach-

[7] See Figure A.2 in the Appendix.

ers), there are a number of additional questions including, for example, information on "intermediate" outcomes. Based on our theoretical framework in Section 2, we are particularly interested in questions that provide information on teacher shortages, teacher motivation, teacher effort, and student and parent (i.e. household) effort. Unlike the management decision tree, additional indicators that measure intermediate outcomes of interest can (and sometimes should) be subjective. For example, if a researcher is interested in whether better management in schools is linked with students' mental health, it would be useful to search the student questionnaires for questions about how students feel in the school (safe, happy, bullied etc.), which are inherently subjective responses. We detail the questions that we classify under each of the intermediate outcome indices in the Appendix and supplementary replication files.

## 3.2 Construction of indices: two examples

We run this exercise using PISA 2012 (OECD, 2021) and Prova Brasil 2013 (INEP, 2021).[8] In total, PISA 2012 has 55 questions that pass the first three decision points, and 49 of them have counterpart topics in the WMS (pass to the fourth decision point). For Prova Brasil 2013, we identify 28 questions and they all have counterpart topics in the WMS.

For each question, we code the responses to fall between 0 and 1, as in the MOPS Census questionnaire, with responses closer to 0 indicating "less structure" – that is, issues are dealt in ad-hoc ways – and responses closer to 1 indicating "more structure" – that is, a set actions that are followed routinely for specific issues. For example, Question 30 from PISA 2012 has the following options:

> Example 4: *"Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviors in your school during the 2011-2012 academic year."*

(t) *I review work produced by students when evaluating classroom instruction.*

[8]For PISA, this also coincides with the "best" year for management questions in the principal survey. In PISA 2015 and 2018 a set of important questions were moved to voluntary teacher questionnaires and, as few countries opted to include the voluntary questionnaires, they are not broadly available. PISA 2022 has returned a few of these important questions to the principal questionnaire resulting in 33 "common" questions. As such, we are able to use 2022 as an additional check on these common questions across time. See Figure A.1. Further, our index is distinct from the "leadership and management" measure from 2012 PISA. The PISA-built index is based off a section of the questionnaire that was titled *management* and contained only a narrow subset of questions. This PISA measure does not compare well to the (empirically robust) management index derived from the World Management Survey (see Liberto et al. (2015)).

With the following options: *(i) did not occur, (ii) 1-2 times during the year, (iii) 3-4 times during the year, (iv) once a month, (v) once a week, (vi) more than once a week.*

In this example, we assign a score of 0 for "did not occur" responses, a score of 1 for "more than once a week", and an equal gradient in between: 0.2 for 1-2 times per year, 0.4 for 3-4 times per year, 0.6 for once a month and 0.8 for once a week. Of note, this is simply measuring the level of structure (here, in terms of frequency) that this principal has in how they conduct their classroom evaluations. The measure at this point is positive rather than normative – whether such frequency (or, "higher structure") is correlated with student outcomes is an empirical question later tested in the validation section.

To construct the indices, we conduct this exercise with every question identified as relevant, then average the scores within each of the practice topics, and build our management index using the same approach as the WMS: (i) standardize each topic score (within country), take the average across the topics and standardize again.[9]

We present below the validation exercises for the set of questions that we can find a benchmark analogue in the WMS, but include robustness to alternative index building tools the Appendix. This method can be replicated using numerous other surveys, as there are many countries that conduct similar national surveys in addition to administering standardized tests across grades. Latin America, for example, is particularly prolific: in addition to Brazil's Prova Brasil, Colombia's SABER, Chile's SIMCE, and Peru's ECE are all available to researchers to conduct a similar exercise to the one detailed here. Further, researchers looking to include measures of management practices in their projects who may not have the budget to run a full WMS-style survey could borrow relevant questions from these existing questionnaires (c.f. Crawfurd et al., 2024).

## 3.3 Validation of new management indices

A key question asked when considering the validity of a new index is whether it is measuring anything of substance relative to outcomes we care about. To test this, we conduct a basic check of the correlation between our management index and student learning outcomes. For each country, we separate schools into quartiles of the management index and show, for each quartile, the average subject test scores in deviations from the country mean. Panel

[9] As a robustness check, Appendix A provides results for alternative index building approaches including principal component analysis and the Anderson (2008) index. The mode of building the index does not have a large impact in the key patterns.

(a) of Figure 2 shows the results for the cross-country averages using PISA's 65 countries. Students in schools in the bottom quartile of (within-country) management score are, on average, performing about 1.5 to 2.6 points lower relative to their own country mean. In contrast, students in schools in the top quartile of (within-country) management score are, on average, performing about 1.8 to 2.4 points higher than their country's mean. To put this into context, 40 PISA points are the equivalent of an average year of learning (OECD, 2019). The range of our results mirror how much, for example, the UK average science score changed between 2009 and 2015 (5 points), and how much the Brazilian average science score decreased over the same period (4 points). We repeat the exercise with data from Prova Brasil in Panel (b), though we restrict the sample to only grade 9 students in 2013 (since PISA focuses on 15-year olds) and standardize math and Portuguese grades within the year. Students in the bottom quartile of the management index have scores that are about 0.4 standard deviations lower, and those in the top quartile have scores that are 0.23 standard deviations higher.

In Table 1 we formalize these relationships by reporting the average correlations between our PISA-based management index and student test scores in reading (Columns 1 to 3), math (Columns 4 to 6) and science (Columns 7 to 9).[10] We report the standard errors in parentheses and *p*-values in square brackets. The standard errors are clustered at the school level and use the appropriate survey weights.[11] In these PISA specifications, we include country fixed effects, and successively introduce school controls (a dummy for private school, dummies for school location, student-teacher ratio, log of the number of students, ratio of computers connected to the web used as a proxy for school resources, and share of government funding relative to total funding the school receives) and then student controls (gender, grade, socio-economic status and immigration status). The top panel includes all schools, and the bottom panel includes schools in Brazil for comparison with the Prova Brasil data. Sample sizes (of the number of students and schools) and the R-squared are reported within each panel.

Column (1) shows the raw relationship between the PISA-based school management index and student performance, only controlling for country fixed effects. The coefficient for all 65 countries is 3.484 points, and for Brazil is 7.410 points. PISA is standardized across years and countries such that the mean is 500 and the standard deviation is 100. As 40 points on the PISA scale is equivalent to one year of learning, the correlation in Column

[10]For these estimates, we use the student-level PISA 2012 dataset and the OECD's `repest` Stata command, which uses the five available test score plausible values for each student and subject.

[11]See Jerrim et al. (2017) for a thorough review of how to best use PISA scores and survey weights.

(1) in the top panel indicates a one standard deviation increase on our management index is associated with higher PISA reading test score points equivalent to about one month's worth of learning. For Brazil, this is equivalent to almost two months. Columns (4) and (7) report similar relationships for math and science scores. Columns (2), (5) and (8) include school controls, which absorb some of the variation and coefficients in the global regression decline by about one point in each of the subjects (about a week's worth of learning). The indicator of private school is large and important in the global regressions, but especially in the Brazilian sample. The coefficient on management practices in the Brazilian sample fall by about half, suggesting that school characteristics and local context absorb a large portion of the variation originally picked up in the management index. Columns (3), (6) and (9) report the fully-specified regression including student controls. Individual student characteristics further reduce the management coefficient but only slightly in the global regressions in Panel A. For Brazil there is a reduction of about 2 points for each subject, again suggesting the composition of students and their characteristics absorbs some of the variation from the management index.

In Table 2 we repeat the exercise with the Prova Brasil-based management index and student scores in Portuguese (Columns 1 to 5) and math (Columns 6 to 10).[12] Here we use the full student-level dataset between 2007 and 2017 (6 rounds), for both grades 5 and 9, and run standard OLS regressions clustering standard errors at the school level. We use the standardized management index and standardized scores for Portuguese and math. Columns (1) and (2) add year and state fixed effects, respectively. Column (3) adds the set of controls that matches those found in the PISA dataset (school controls: dummies for school location, student-teacher ratio, log of the number of students, and dummies for a computer lab and for internet access; student controls: gender, socio-economic status, and race). Column (4) includes additional controls available in the Prova Brasil data (school controls: dummies for principal age, education, race, and other employment, composition of gender and race of teachers, and teachers holding a college degree, average teacher tenure; student control: dummies for mother's education). Finally, Column (5) adds school fixed effects (and drops state fixed effects), which allows us to compare school changes in learning results with their changes in management practices. In this more demanding specification,

[12] Unlike PISA, the Prova Brasil dataset includes school identifiers that allow for a one-to-one match with the schools surveyed for the WMS. We are able to match 273 schools in the 2013 waves of both surveys. Figure A.3 shows a binned scatter plot of the WMS management score against the standardized Prova Brasil-based management score for these 273 schools across the same set of questions. Each circle represents the average of five schools. There is a positive and significant correlation, suggesting reasonable internal validation of the Prova Brasil index.

the coefficient suggests that one standard deviation higher management score is associated with a 0.017 to 0.019 standard deviation higher score in Portuguese and math.[13]

Taken together, these results suggest the management indices we built from these two distinct datasets yield measures that are correlated with student outcomes. The relationships hold even after the inclusion of standard controls in education regressions, providing evidence that the index is measuring something unique and important about the processes at these schools.

## 3.4 Advantages and limitations

We chose these two public datasets as they illustrate different strengths and weaknesses: PISA provides a global view and includes a sample of both public and private schools, while Prova Brasil is a public sector census that contains school identifiers enabling matching with external datasets. Both, however, share the downside that the data are self-reported and collected by external parties. The WMS, on the other hand, has detailed and independent measurement and can be adjusted to fit the researchers' needs, but is expensive and lengthy to collect. We expand on these issues below.

**Issues with self-reported data.** One concern with self-reported data is measurement equivalence. To address potential measurement error driven by cross-cultural understandings and norms in answering questions, we standardize our PISA-based management index *within countries*. This has an important implication: since all 65 countries have a mean score of zero, our index cannot be used to construct cross-country rankings of school management. Instead, the value of our PISA-based index lies in enabling academics and practitioners to study the (within-country) relationship between management and other variables for a far wider set of countries than was previously possible. This issue of cross-cultural norms is less of a concern for our Prova Brasil-based index since it is a national dataset.

Another concern with self-reported data is that it is difficult to assess whether respondents are being accurate and truthful. The WMS methodology includes strategies to elicit truthful information during the interview (such as always asking open-ended questions and asking for examples), but these are not available in self-reported questionnaires. We partially address this issue by focusing on objective questions and, in this first 'proof of concept', also on the topics that have a direct equivalent in the WMS to allow for benchmarking.

[13] When exploring the longitudinal dimension of the WMS data across multiple countries and adding firm fixed effects to their management-productivity regressions, Bloom et al. (2012) also find a weaker relationship. Their coefficient is 0.047 standard deviation higher productivity.

**Issues with question wording and permanence.** While public datasets provided by national or international organizations are free, they are also entirely governed by those institutions. This means that the wording of questions can (and often does) change between waves, or questions are dropped/re-instated as the organization sees fit. This can create issues for longitudinal analysis or if researchers needed a specific year that does not have enough relevant questions. The issue of missing questions is more conceptual: ultimately the index will be a measure of the available information, and if there are fewer questions available the measure will be based on a narrower set of information. Whether it is enough to proxy for the construct that the researcher is looking to measure is an individual conceptual decision, and to some extent an empirical question.

**Issues with missing values.** A benefit of large public datasets is that they often have a wealth of additional data beyond the narrow scope of a researcher-specific field survey. However, sometimes these large datasets also have many missing values. There are a number of standard methods for handling missing data; our preferred method when missing data is in the set of control variables is to impute a value outside the support of the variable distribution (the number 99 is a common choice) and include a dummy variable that identifies the values which were imputed. When data is missing for critical variables (say, outcome variables or key explanatory variables), we drop the observation.

**Issues with the WMS.** There are a number of limitations in surveys like the WMS, discussed in detail in Scur et al. (2021) and Bloom et al. (2016).[14] The most important considerations in the education surveys are (i) the high cost and (ii) limited set of topics. On (i), the cost of carrying out of a full survey can be prohibitive for some projects, especially those run (and funded) by graduate students or projects in developing countries. It is not clear that the WMS is the best option for such projects, and we hope our method provides a useful alternative. On (ii), the set of practices measured in the WMS are certainly not exhaustive and by design mostly covers basic practices used across a number of sectors. While researchers can somewhat adjust the survey to their needs (i.e. add questions) it is already long and has limited space for additional questions without removing existing ones. Depending on the number of questions removed, this could impact the comparability of the remaining questions and take away this key benefit of the WMS vs other surveys. Further, adding questions can be costly if done through the appropriate process including piloting and psychometric validation. Questions that appear in public surveys (such as PISA) have

[14] See for example Table 1 in Bloom et al. (2016), reproduced in the Appendix, for a succinct summary.

already gone through these processes and are ready for use.

## 4 Applications: unpacking principals and practices

While this paper primarily focuses on management practices and structures, all practices are at the outset implemented and adopted by *people*.[15] The way we conceive of management practices is as described in Quinn and Scur (2021): as a "kind of supporting scaffold structure in which managers operate". In their paper, they propose that successful outcomes in "a [school] with very few formal management processes in place [...] where teachers arrive, go into their classrooms, and start teaching the content of the day" will heavily depend on whether the principal is effective or not. In this scenario, quality will be strongly linked with individual characteristics of the principal, and the whole school community – students, parents, teachers – will depend on the star principal remaining at the school (or, alternatively, a bad principal leaving the school). As they note,

> "Without a solid scaffolding, much is left to chance. But consider instead a [...] school that has a minimal set of formal processes in place, such that there is a basic level of organizational scaffolding to structure workers daily activities. [...] When they arrive there is less left to chance: such workers have clear parameters of what they are expected to achieve and how. If this organization has a very poor manager, the workers will still have at least a basic clarity on what they are expected to achieve, and how their performance will be assessed; in many respects, this provides an effective 'floor' to the organization's performance."

In short, while people are the ones who adopt and implement practices, there exists a level of formality to practices that give way to some level of continuity and *permanence*. In the WMS scoring rubric, a management practice scores more highly if it is *structured*—i.e., it has been written down, or codified in some way, rather than simply being embodied in the behaviour of the school principal. Conceptually, principal turnover could be good (if an incumbent principal is ineffective) or bad (if an incumbent principal is effective). With the public datasets at our disposal we cannot explore this nuance as typically these datasets do not have principal identifiers to allow for estimation of principal fixed effects. However, researchers with access to such data alongside our new indices could directly explore the association between principal quality and management.

[15] Huber et al. (2021b), for example, show that firms in Germany incurred significant losses when Jewish managers were forced out as a result of rising discrimination in Nazi Germany.

We expect that both principals *and* management practices matter, but how much each contribute to student learning is an empirical question that is now an open field with our new indices. In this section we present exploratory evidence with public data that can start to move the needle on this question. In Section 4.1 we document the relationship between managers, management practices and student learning both on average and by exploiting unexpected principal turnover events. We show that managers matter, but so do managerial structures. In Section 4.2 we turn back to the question of *how* these structures drive student learning, and document empirical support for the theoretical predictions in Section 2.

## 4.1 Principals and practices: evidence from principal turnover

A natural setting to study the relationship between principals and practices is when principals unexpectedly change. Akhtari et al. (2022) study mayoral elections in 2008 and 2012 in Brazil and find that when cities get a new government following close elections there is an "upheaval" in the municipal bureaucracy, including school principals. They note that there is an "increase in the replacement rate of personnel in schools controlled by the municipal government" and that this is accompanied by "test scores that are 0.05 to 0.08 standard deviations lower. In contrast, turnover of the mayor's party does not impact local (non-municipal) schools." They argue that changes in political leadership trigger changes in school personnel (both school principals and teachers) which, in turn, negatively affect test scores. But why is this change in personnel so detrimental? What is the channel this effect is working through? We use our Prova Brasil-based management index to further explore this result, and, specifically, ask does political turnover negatively impact student test scores *through a school management channel*?

To do this, we merge our Prova Brasil management index into the Akhtari et al. (2022) replication dataset (Moreira et al., 2021). Figure 3 replicates their main regression discontinuity design comparing outcomes in municipalities where the incumbent mayor *narrowly won* with those where they *narrowly lost*, under the assumption that turnover in political party is as good as random in these close-election municipalities.[16] In Panel (a) we directly replicate their result on student test scores: "party turnover lowers test scores, as measured one year after the election, by 0.05 to 0.08 standard deviation units in terms of the individual-level national distribution of test scores." They found the result to be persistent up to five years after political turnover, and the magnitude is substantial: about one-third of the standard deviation-level impact of some of the most successful educational improvement interventions.

[16] They find empirical support for this identification assumption. We do not replicate all aspects of their work here; for more details see Akhtari et al. (2022).

Panel (b) replicates their placebo test, where they consider whether state-run public schools in the same municipality as the municipal-run public schools had any change in test scores.

Using the same strategy, we replace the outcome variable of test scores with our management index in Panels (c) and (d). There is a clear discontinuity in both test scores and management scores in municipal schools: both are *lower* in cities where the incumbent mayor narrowly loses an election (right side of the graph) than in cities where the incumbent mayor narrowly wins (left side of the graph). Panel (d) shows that this discontinuity is not present in non-municipal schools run by the state (rather than municipal) government. This is an important placebo test since these schools would not have been subject to "upheaval" associated with political turnover at the municipal level. Table A.2 shows the associated regression analysis for the results in Figure 3. Akhtari et al. (2022) argue that changes in political leadership trigger changes in school personnel and that this translates into lower test scores. We show that the channel this is likely working through is the decline in the quality of management at schools in those municipalities.

### 4.1.1 What is driving the decline in management practices?

We show that the impact is driven by municipalities where there is principal turnover (Figure 4a) and not when principals remain in place (Figure 4b). While this sub-sample may depart from the clean quasi-experiment if, say, the decision to change principals is related to principal characteristics, we can turn to a particular institutional arrangement of Brazilian municipalities for further support. Specifically, Brazilian municipalities differ in that some have *politically appointed public school staff* while others fill these positions via civil service selection exams. Since the latter group are plausibly less susceptible to political "upheaval", this offers a relatively more exogenous variation in principal turnover. The discontinuity in school management scores is present in the group of municipalities with an appointed structure (Figure 4c), and not where principals are civil servants (Figure 4d). Consistent with management practices being an important channel in explaining changes in student test scores, we find that test scores are indeed more significantly affected in municipalities with principal turnover and political appointment structures, relative to those with no "upheaval" (Appendix Figure A.4).

### 4.1.2 Practices or characteristics?

In our concept of management as a "scaffold", schools that have reached a minimum level of structured practices should be less affected by upheaval in the bureaucracy and school staff changes. Using the full dataset, we undertake a correlational analysis to consider the broader relationship between turnover, principal characteristics, management structures and student outcomes. Specifically, we run the following regression:

$$TestScore_{igst} = \alpha + \beta_1 M_{st} + \beta_2 NewPrincipal_{st} + \gamma_1 X_{st} + \gamma_2 P_{st} + \eta_r + \delta_t + \varepsilon_{ict} \quad (2)$$

where $TestScore_{igst}$ is the standardized test score in Portuguese or math for student $i$ in grade $g$ in school $s$ at time $t$, $M_{st}$ is Prova Brasil management index for school $s$ at time $t$, $NewPrincipal_{st}$ is an indicator for whether the principal is new at school $s$ in year $t$, $X_{st}$ include school, teacher and student controls, $P_{st}$ include principal characteristics, $\eta_r$ are state fixed effects and $\delta_t$ are year fixed effects. School controls include: whether the school is in an urban area, the student-teacher ratio, school size (log number of students), whether the school has a computer lab an internet access and the level of spending in education in the school's municipality. Teacher controls include: gender, race, education level and tenure at the school. Student controls include: age, gender, socioeconomic status, race, parent literacy and whether the student has a job. Principal characteristics include: gender, age, seniority, whether the principal has a college degree, whether the degree is in pedagogy, and wage band level. Where characteristics are categorical variables in the data, we build indicators that take a value of 1 when categories are above median.

Table 3 shows the results; columns (1) to (5) with Portuguese scores as the outcome and columns (6) to (10) with math scores as the outcome. As Table 2 iterates over sets of controls, all columns here already include year and state fixed effects as well as all student, teacher and school controls. Columns (1) and (6) show that, conditional on all controls, one standard deviation higher management score is correlated with 0.033 standard deviations higher Portuguese and 0.037 higher math test scores, respectively.[17] Columns (2) and (7) show evidence that principal turnover is generally detrimental for student performance: conditional on controls, a new principal is correlated with a decline of 0.039 in Portuguese scores and 0.043 math scores. Columns (3) and (8) include a set of principal characteristics. Male, older principals and those with over 15 years of service are correlated with lower achievement in both subjects. Students in schools whose principal has higher education, higher wages, and who only work in one school tend to do better.

Columns (4) and (9) include the management index and turnover indicator. While the

[17] This is effectively the same specification as Columns (4) and (9) in Table 2. The controls are slightly different because the purpose of Table 2 is to build a closer match with PISA, whereas our goal in the present analysis is to include all relevant controls available in the dataset. Reassuringly, they are very similar.

coefficient on the management index barely moves, the coefficient on turnover changes substantially. Columns (5) and (10) include all variables together. Of note, the size of the coefficients of principal characteristics are relatively similar when controlling for the management index and turnover, suggesting even after controlling for management practices principal characteristics still matter. The management index is also broadly similar (consistent with our theoretical framework), but the turnover variable about halves after controlling for both management practices and manager characteristics. The evidence from Akhtari et al. (2022) shows that when there is turnover in municipal administrations and schools, student test scores decline. One interpretation of the evidence in Table 3 is that the deleterious impact of the principal turnover is lessened when controlling for management structures.

Taken together, these results show a simple but powerful application of our measurement approach. We took existing public data from Prova Brasil and linked them to data in the American Economic Review replication archives. With no further survey costs, we were able to explore mechanisms around how political turnover affects student learning outcomes, and add novel evidence to the question of how much management matters relative to managers in schools. Our findings suggest that both principal characteristics and management practices are important in attenuating the negative impact of principal turnover; in fact, the negative coefficient on "new principal" falls by half between Columns 2 and 5 (7, 10), and is equally attributable to both factors (see Columns 4, 9).

## 4.2 How management practices matter

Good management practices are correlated with student learning across a wide range of studies, and we show that it may also mitigate the negative impact of principal turnover. But *how* do simple structures have such an important impact on student learning? Our theoretical framework set out in in Section 2 outlines channels through which management may matter for student outcomes. These channels are distinct from principal characteristics and leadership, and focus on teachers, students and parents. Our data includes rich information on intermediate outcomes for these actors, which we use to test the predictions of our model.[18]

**Prediction 1: Teacher shortages.** In Table 4 columns (1) and (2), the dependent variable is the teacher shortage index and the explanatory variable of interest is the school

[18]Exploring the effects of principal turnover in this setting, both theoretically and empirically, is outside the scope of this paper as one would need data on principal quality to do so. We believe this is a fruitful area of further research.

management index. In both panels, these indices are standardised within-country. In Panel A, for PISA, Column (1) includes only country fixed effects. Column (2) adds school and student controls. Consistent with the theoretical prediction, there is a negative relationship between management and teacher shortage: a one standard deviation increase in the management index is associated with a 0.057 standard deviation decrease in the teacher shortage index. In Panel B, for Prova Brasil, Column (1) includes only year effects. Column (2) adds PISA-like controls, additional Prova Brasil controls, and school fixed effects. Again, there is a negative relationship: a one standard deviation increase in the school management index is associated with a 0.088 standard deviation decrease in the teacher shortage index.

**Prediction 2: Teacher motivation.** In Table 4 columns (3) and (4) report the specifications with our teacher motivation index as the dependent variable. As predicted, there is a positive relationship in both panels. In Panel A, for PISA, a one standard deviation increase in the school management index is associated with a 0.316 standard deviation increase in the teacher motivation index. In Panel B, for Prova Brasil, a one standard deviation increase in the school management index is associated with a 0.218 standard deviation increase in the teacher motivation index.

**Prediction 3: Teacher effort.** In Table 4 columns (5) and (6) report the results with the teacher effort index as the dependent variable. Again, there is a positive relationship. In Panel A, for PISA, a one standard deviation increase in the school management index is associated with a 0.071 standard deviation increase in the teacher effort index. In Panel B, for Prova Brasil, a one standard deviation increase in the school management index is associated with a 0.059 standard deviation increase in the teacher effort index.

**Prediction 4: Household effort.** In Table 4 columns (7) and (8) we report results with the household effort index as the dependent variable. Once again there is a positive relationship. In Panel A, for PISA, a one standard deviation increase in the school management index is associated with a 0.282 standard deviation increase in the household effort index. In Panel B, for Prova Brasil, a one standard deviation increase in the school management index is associated with a 0.054 standard deviation increase in the household effort index.

These findings suggest that the causal pathways from the quality of management practices to student learning posited in the theory — selection and incentives within and beyond the school — are empirically plausible across a wide range of countries.

# 5 Conclusion

Policy makers have begun to set ambitious, universal learning goals. To achieve these targets it will be necessary to understand why — within and across current education systems — some students are learning more in some schools than others. Although there are likely many factors at work, at least part of this variation in learning stems from the management of schools. To explore this issue, researchers and practitioners need to be able to measure school management accurately and cost-effectively at scale across schools and countries, and be in a position to postulate mechanisms behind any observed relationship between school management and student learning.

We make three key contributions with this paper. The first contribution is the development of a new approach to measurement at scale using existing public data sources. Both of our new school management indices confirm the strong positive correlation of school management scores with school-level student outcomes first reported in Bloom et al. (2015). The second contribution is a theoretical framework to explore the potential channels via which management structures impact student outcomes. We propose a specific way they may work through selection and incentives of the actors within the organization: teachers, students and parents. The third contribution is documenting new patterns with our management indices that provide evidence for the role that *both* managers and management practices play in supporting student learning.

An important implication from our results is that if researchers look only at the effect of principal-specific attributes, the estimated relationship is likely to bundle both the principal's managerial prowess (leadership, charisma, etc.) as well as the practices that they implemented in their school. Our evidence suggests that un-bundling these is important, as *both* contribute to student performance, but management practices could be relatively more actionable as a policy lever. Charisma and leadership may be difficult to teach but recent research suggests it can be effective to teach managers how to adopt better management practices (c.f. Anand et al., 2023; Beg et al., 2023; Bloom et al., 2013; Roland G. Fryer, 2017).

# References


Adelman, M. and R. Lemos (2021). *Managing for Learning: Measuring and Strengthening Education Management in Latin America and the Caribbean.* International Development in Focus. Washington, DC: World Bank.

Akhtari, M., D. Moreira, and L. Trucco (2022, February). Political Turnover, Bureaucratic Turnover, and the Quality of Public Services. *American Economic Review 112*(2), 442–93.

Akmal, M. and L. Pritchett (2019). Learning equity requires more than equality: Learning goals and achievement gaps between the rich and the poor in five developing countries. Working Paper 19/028, RISE.

Anand, G., A. Atluri, L. Crawfurd, T. Pugatch, and K. Sheth (2023). Improving school management in low and middle income countries: A systematic review. *Economics of Education Review 97*(C).

Anderson, M. L. (2008). 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*(484), 1481–1495.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020). CEO behavior and firm performance. *Journal of Political Economy 128*(4), 1325–1369.

Beg, S. A., A. E. Fitzpatrick, and A. Lucas (2023, October). Managing to learn. Working Paper 31757, National Bureau of Economic Research.

Bender, S., N. Bloom, D. Card, J. Van Reenen, and S. Wolter (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics 36*(S1), S371–S409.

Bertrand, M. and A. Schoar (2003). Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics 118*(4), 1169–1208.

Biasi, B. (2021, August). The labour market for teachers under different pay schemes. *American Economic Journal: Economic Policy 13*(3), 63–102.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2019, May). What drives differences in management practices? *American Economic Review 109*(5), 1648–83.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does Management Matter? Evidence from India. *The Quarterly Journal of Economics 128*, 1–51.

Bloom, N., C. Genakos, R. Sadun, and J. Van Reenen (2012). Management practices across firms and countries. *Academy of Management Perspectives 26*(1), 12–33.

Bloom, N., R. Lemos, R. Sadun, D. Scur, and J. Van Reenen (2016, May). International data on measuring management practices. *American Economic Review 106*(5), 15256.

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

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen (2017, September). Healthy business? Managerial education and management in healthcare. Working Paper 23880, National Bureau of Economic Research.

Bloom, N., C. Propper, S. Seiler, and J. V. Reenen (2015). The impact of competition on management quality: Evidence from public hospitals. *Review of Economic Studies 82*(2), 457–489.

Bloom, N., R. Sadun, R. Lemos, and J. Van Reenen (2019). Healthy business? Managerial education and management in healthcare. *The Review of Economics and Statistics 0*(ja), 1–45.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics 122*, 1351–1408.

Buffington, C., L. Foster, R. Jarmin, and S. Ohlmacher (2017). The management and organizational practices survey: An overview. *Journal of Economic and Social Management 42*(1), 1–26.

Böhlmark, A., E. Grönqvist, and J. Vlachos (2016, October). The Headmaster Ritual: The Importance of Management for School Outcomes. *Scandinavian Journal of Economics 118*(4), 912–940.

Cornwell, C., I. Schmutte, and D. Scur (2021). Building a productive workforce: the role of structured management practices. *Management Science 67*(12).

Crawfurd, L. (2017, 08). School Management and PublicPrivate Partnerships in Uganda. *Journal of African Economies 26*(5), 539–560.

Crawfurd, L., T. Pugatch, and K. Sheth (2024, April). Developing school leadership at scale: Experimental evidence from schools for marginalized communities in India. AEA RCT Registry.

Cullen, J. B., S. D. Levitt, E. Robertson, and S. Sadoff (2013, May). What can be done to improve struggling high schools? *Journal of Economic Perspectives 27*(2), 133–52.

Dal Bó, E. and F. Finan (2020). At the intersection: A review of institutions in economic development. In J.-M. Baland, F. Bourguignon, J.-P. Platteau, and T. Verdier (Eds.), *The Handbook of Economic Development and Institutions*, Chapter 1. Princeton and Oxford: Princeton University Press.

Delfgaauw, J., R. Dur, C. Propper, and S. Smith (2011, 07). Management practices: Are not for profits different? *SSRN Electronic Journal*.

Dohmen, T. and A. Falk (2010, May). You get what you pay for: Incentives and selection in the education system. *The Economic Journal 120*, 1–27.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica 90*(3), 1063–1084.

Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Volume 2 of *Handbook of Economic Field Experiments*, pp. 467 – 514. North-Holland.

Fryer, R. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics 129*(3), 1355–1407.

Fryer, R. (2017, May). Management and student achievement: Evidence from a randomized field experiment. Working Paper Series 23437, NBER.

Giorcelli, M. (2019, January). The long-term effects of management and technology transfers. *American Economic Review 109*(1), 121–52.

Hanushek, E. A., S. Link, and L. Wössmann (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics 104*(C), 212–232.

Hanushek, E. A. and L. Wössmann (2006, March). Does educational tracking affect performance and inequality? Differences- in-differences evidence across countries. *The Economic Journal 116*(510), 63–76.

Huber, K., V. Lindenthal, and F. Waldinger (2021a). Discrimination, managers, and firm performance: Evidence from aryanizations in nazi germany. *Journal of Political Economy 129*(9), 2455–2503.

Huber, K., V. Lindenthal, and F. Waldinger (2021b). Discrimination, Managers, and Firm Performance: Evidence from Aryanizations in Nazi Germany. *Journal of Political Economy 129*(9), 2455–2503.

INEP (2021). Microdados Prova Brasil (SAEB). https://dados.gov.br/dataset/inep-microdados-do-saeb. Accessed: 2021-10-22.

Jerrim, J., L. A. Lopez-Agudo, O. D. Marcenaro-Gutierrez, and N. Shure (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review 61*(C), 51–58.

Kala, N. (2019, September). The impacts of managerial autonomy on firm outcomes. Working Paper 26304, National Bureau of Economic Research.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal 125*(588), F397–F424.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review 10*(3), 179–214.

Leaver, C., O. Ozier, P. Serneels, and A. Zeitlin (2021, July). Recruitment, effort and retention effects of performance contracts for civil servants: Experimental evidence from rwandan primary schools. *American Economic Review 111*(7), 2213–46.

Lemos, R., K. Muralidharan, and D. Scur (2021, January). Personnel management and school productivity: Evidence from india. Working Paper 28336, National Bureau of Economic Research.

Lemos, R., K. Muralidharan, and D. Scur (2024, 04). Personnel Management and School Productivity: Evidence from India. *The Economic Journal 134*(661), 2071–2100.

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

McConnell, K. J., K. A. Hoffman, A. Quanbeck, and D. McCarty (2009). Management practices in substance abuse treatment programs. *Journal of Substance Abuse Treatment 37*(1), 79–89.

McCormack, J., C. Propper, and S. Smith (2014). Herding cats? Management and university performance. *The Economic Journal 124*(578), F534–F564.

Moreira, D., M. Akhtari, and L. Trucco (2021). Data and Code for: Political Turnover, Bureaucratic Turnover and the Quality of Public Services. https://doi.org/10.3886/E150323V1. Accessed: 2021-10-22.

OECD (2019). *How PISA results are reported: What is a PISA score?*

OECD (2021). PISA Database. https://www.oecd.org/pisa/data/. Accessed: 2021-10-22.

Prem, M. and P. Munoz (2023). Managers Productivity and Recruitment in the Public Sector: The Case of School Principals. *American Economic Journal: Economic Policy forthcoming*.

Quinn, S. and D. Scur (2021, 06). Management practices and public policy: an overview. *Oxford Review of Economic Policy 37*(2), 221–230.

Rasul, I. and D. Rogger (2016). Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service. *The Economic Journal 128*(608), 413–446.

Roland G. Fryer, J. (2017, May). Management and Student Achievement: Evidence from a Randomized Field Experiment. NBER Working Papers 23437, National Bureau of Economic Research, Inc.

Romero, M., J. Sandefur, and W. A. Sandholtz (2020, February). Outsourcing education: Experimental evidence from liberia. *American Economic Review 110*(2), 364–400.

Rothstein, J. (2015). Teacher quality policy when supply matters. *American Economic Review 105*(1), 100–130.

Ruhose, J. and G. Schwerdt (2016). Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries. *Economics of Education Review 52*(C), 134–154.

Scur, D., R. Sadun, J. Van Reenen, R. Lemos, and N. Bloom (2021, 06). The World Management Survey at 18: lessons and the way forward. *Oxford Review of Economic Policy 37*(2), 231–258.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

West, M. R. and L. Wössmann (2010). Every catholic child in a catholic school: Historical resistance to state schooling, contemporary private competition and student achievement across countries. *The Economic Journal 120*(546), F229–F255.

World Bank (2018). *World Development Report 2018: Learning to Realize Educations Promise*. Washington, DC: World Bank.

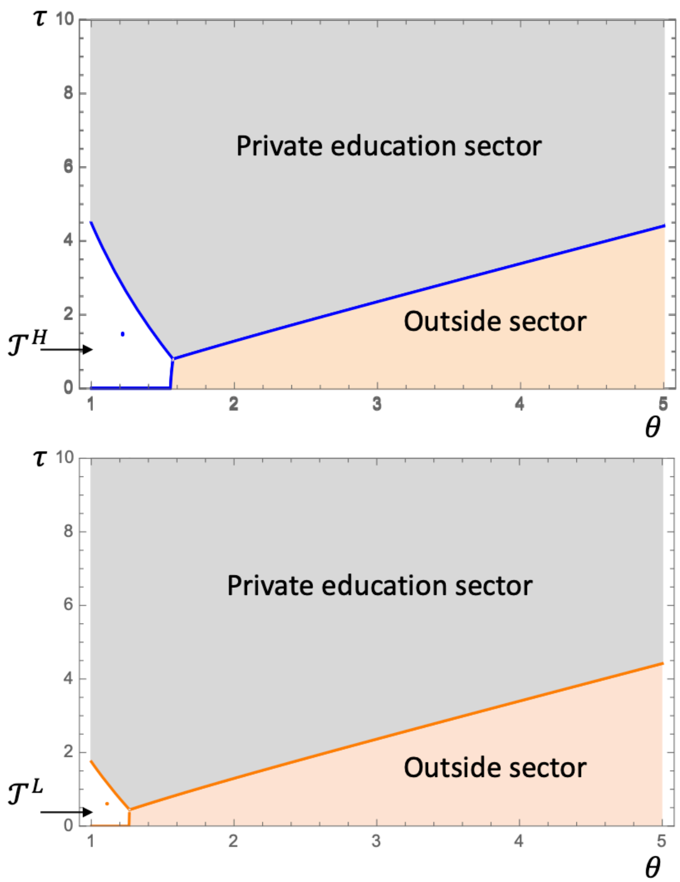
Wössmann, L. (2005). The effect heterogeneity of central examinations: Evidence from TIMSS, TIMSSRepeat and PISA. *Education Economics 13*(2), 143–169.

Wössmann, L. (2016, September). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives 30*(3), 3–32.

Wössmann, L., E. Lüdemann, G. Schütz, and M. R. West (2007). School accountability, autonomy, choice, and the level of student achievement. *OECD Publishing, Paris* (13).

# Figures and Tables

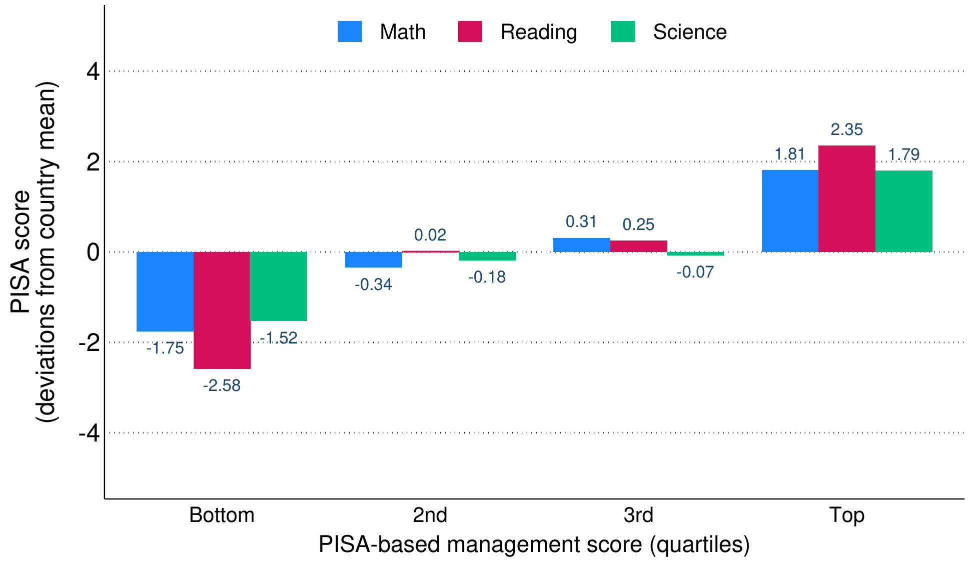
Figure 1: Teacher selection



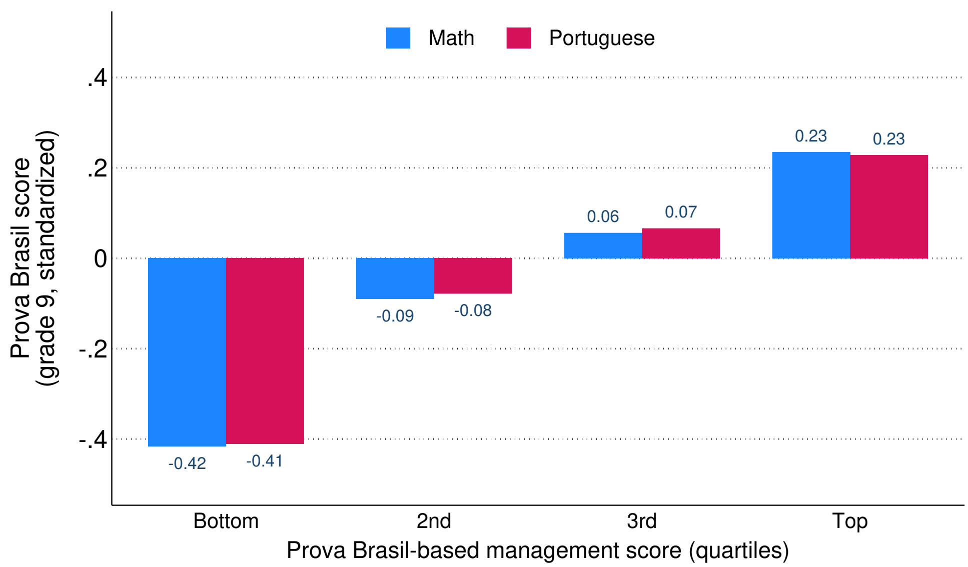

*Note:* Teacher ability is distributed $\theta \sim U\,[1,5]$ and teacher intrinsic motivation is distributed $\tau \sim U\,[0,10]$. In the low management public school $G^L = 30$, $\Delta^L = 0$, and $\gamma^L = 1$. In the high management public school, $G^H = 35$, $\Delta^H = 0.5$ and $\gamma^H = 2$. Other parameters are set at $W = 15$, $B = 40$, $\bar{y} = 4.5$, $\beta = 50$, and $\bar{z} = 1$. The blue point in the top panel shows average teacher ability and average baseline intrinsic motivation among teacher types who select into a high management public school. The $(x, y)$-coordinates are $(1.21, 1.47)$. The orange point in the bottom panel shows average teacher ability and average baseline intrinsic motivation among teacher types who select into a low management public school. The $(x, y)$-coordinates are $(1.11, 0.60)$.

Figure 2: Management Index validation

(a) PISA



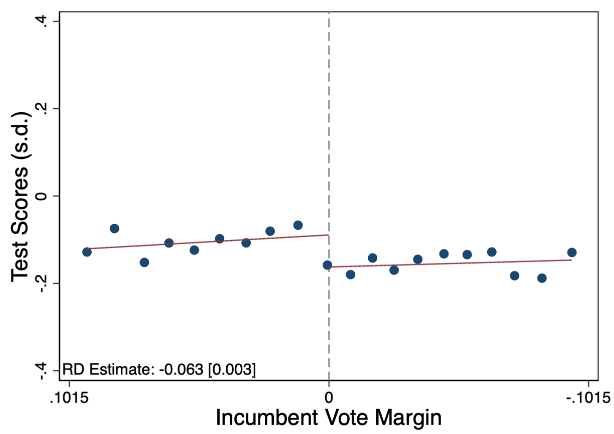

(b) Prova Brasil



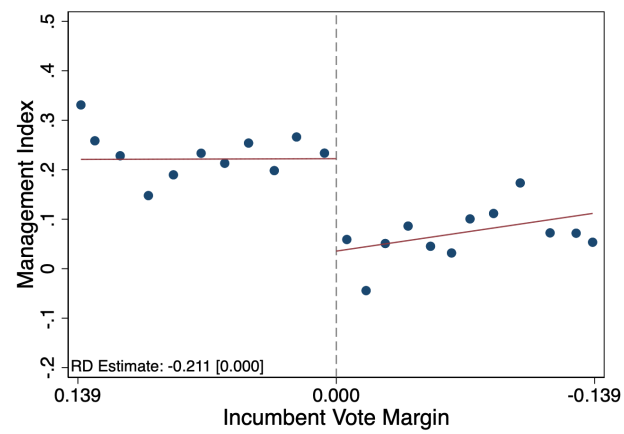

*Note:* Data for Panel (a) is from PISA 2012 including 15,196 schools across 65 countries. Management indices standardized within country. Student outcomes are estimated using five plausible values and collapsed at the school level using PISA's senate weights. Quartiles of the management index are built at the country level. Test scores are presented as deviations from the subject-specific country mean. Data for Panel (b) is from Prova Brasil (2013), including 33,344 schools. Sample restricted to schools with grade 9 to maintain closer comparability to the WMS sample. Quartiles of the management index are built from this sample. Test scores are presented as deviations from the subject-specific mean within the same sample.

Figure 3: Political turnover, test scores, and school management

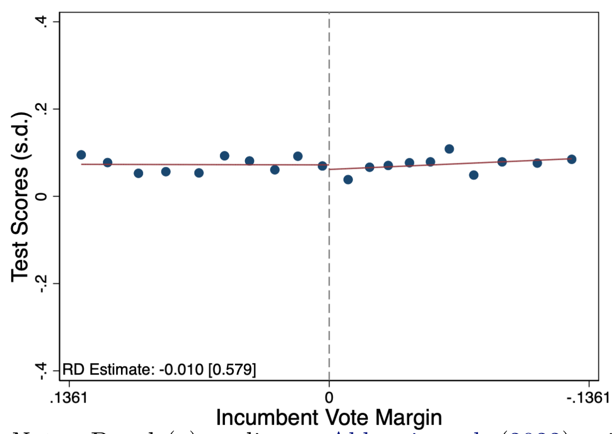
(a) Test Scores: municipal (treated) schools



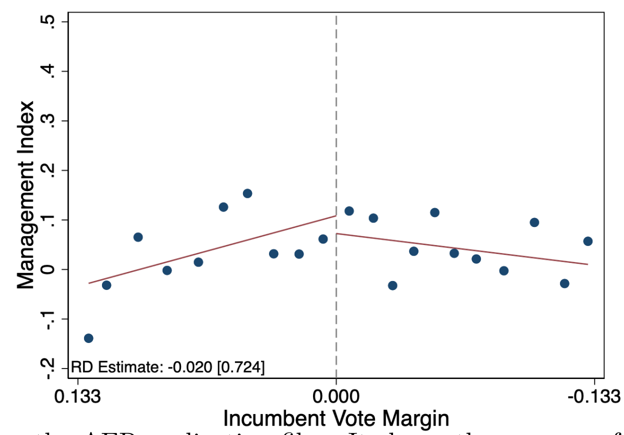

(b) Management: municipal (treated) schools




(c) Test Scores: non-municipal (placebo) schools



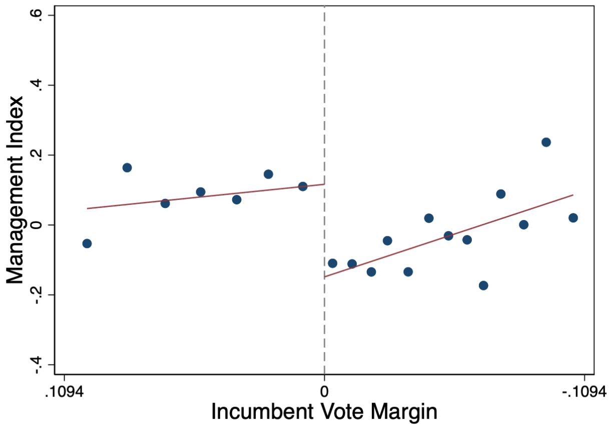

(d) Management: state (placebo) schools



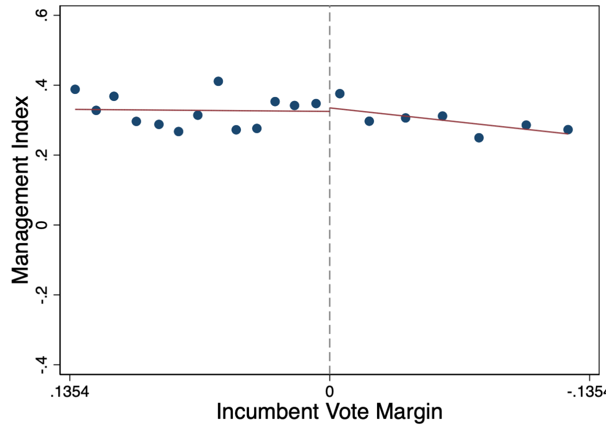

*Note:* Panel (a) replicates Akhtari et al. (2022) using the AER replication files. It shows the average of individual-level test scores by bins of $IncumbVoteMargin$ in municipal schools, pooling students from grade 5 and grade 9 and controlling for the average, school-level test scores for the respective grade at baseline. Municipalities with $IncumbVoteMargin < 0$ experienced a change in the political party of the mayor. Municipalities with $IncumbVoteMargin > 0$ did not experience a change in the political party of the mayor. Note that values to the right side of the zero are negative (political turnover), while values on the left side are positive (no political turnover). Selected bandwidth follows Calonico et al. (2017). Panel (b) repeats the analysis but with the average standardized management score in municipal schools by bins of $IncumbVoteMargin$ as the outcome variable, controlling for the standardized management score in the baseline year (year before the election). Panels (c) and (d) repeat the analysis for for non-municipal schools (a placebo test).

Figure 4: Political turnover, principal turnover, school management and student outcomes

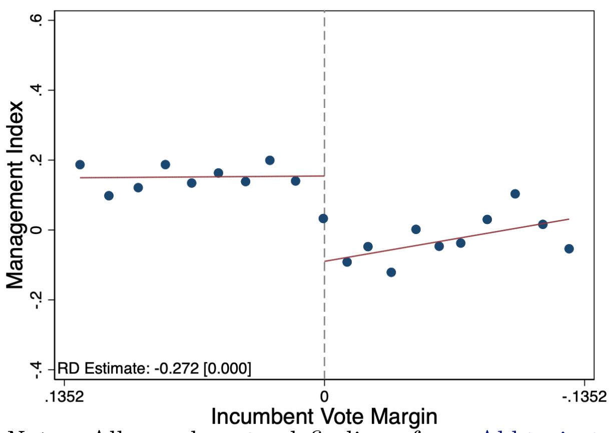
(a) New principal



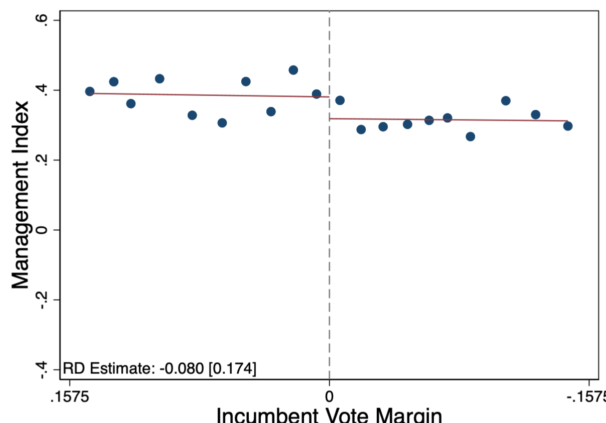

(b) No principal turnover




(c) Principals are politically appointed




(d) Principals are civil servants




*Note:* All panels extend findings from Akhtari et al. (2022) using the AER replication files. It shows the average of average standardized management score by bins of $IncumbVoteMargin$ in municipal schools, controlling for the standardized management score in the baseline year (year before the election). Municipalities with $IncumbVoteMargin < 0$ experienced a change in the political party of the mayor. Municipalities with $IncumbVoteMargin > 0$ did not experience a change in the political party of the mayor. Note that values to the right side of the zero are negative (political turnover), while values on the left side are positive (no political turnover). Selected bandwidth follows Calonico et al. (2017). Panels (a) and (b) restrict the sample to those schools which had a principal turnover and those which did not (respectively). Panels (c) and (d) restrict the sample to schools in municipalities where principals are politically appointed or part of the civil service (respectively).

Table 1: Management and student performance, PISA

| | Reading PISA Points | | | Math PISA Points | | | Science PISA Points | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **All Schools** | | | | | | | | | |
| Management Index | 3.484 | 2.358 | 2.231 | 3.001 | 2.094 | 1.898 | 2.729 | 1.919 | 1.772 |
| | (1.022) | (0.985) | (0.809) | (0.995) | (1.007) | (0.799) | (0.999) | (1.018) | (0.805) |
| | [0.001] | [0.017] | [0.006] | [0.003] | [0.038] | [0.018] | [0.006] | [0.059] | [0.028] |
| Private=1 | | 11.140 | 2.732 | | 11.058 | 1.799 | | 9.844 | 1.073 |
| | | (2.867) | (2.548) | | (2.840) | (2.636) | | (2.737) | (2.369) |
| | | [0.000] | [0.284] | | [0.000] | [0.495] | | [0.000] | [0.650] |
| Students | 410701 | 410200 | 410200 | 410701 | 410200 | 410200 | 410701 | 410200 | 410200 |
| Schools | 15196 | 15176 | 15176 | 15196 | 15176 | 15176 | 15196 | 15176 | 15176 |
| R-Squared | 0.242 | 0.289 | 0.423 | 0.306 | 0.341 | 0.449 | 0.299 | 0.329 | 0.431 |
| **Brazil** | | | | | | | | | |
| Management Index | 7.410 | 3.409 | 1.321 | 8.817 | 4.850 | 2.644 | 9.924 | 5.966 | 4.014 |
| | (2.867) | (2.468) | (1.881) | (2.476) | (2.350) | (1.836) | (2.483) | (2.267) | (1.839) |
| | [0.010] | [0.167] | [0.482] | [0.000] | [0.039] | [0.150] | [0.000] | [0.009] | [0.029] |
| Private=1 | | 40.288 | 31.708 | | 39.771 | 28.989 | | 36.134 | 26.586 |
| | | (16.616) | (13.774) | | (15.540) | (13.267) | | (12.022) | (9.696) |
| | | [0.015] | [0.021] | | [0.010] | [0.029] | | [0.003] | [0.006] |
| Students | 14949 | 14949 | 14949 | 14949 | 14949 | 14949 | 14949 | 14949 | 14949 |
| Schools | 561 | 561 | 561 | 561 | 561 | 561 | 561 | 561 | 561 |
| R-Squared | 0.009 | 0.173 | 0.352 | 0.014 | 0.219 | 0.390 | 0.018 | 0.200 | 0.341 |
| Country FE | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| School Controls | | Y | Y | | Y | Y | | Y | Y |
| Student Controls | | | Y | | | Y | | | Y |

*Note:* Standard errors in parentheses, p-values in square brackets. OLS regressions were run with the student-level PISA dataset using the OECDs `repest` Stata command. Standard errors are clustered at the school level and use all 5 plausible values for each subject and student final weights. Main independent variable is the PISA-based management index standardized using the overall distribution. All specifications include country fixed effects (except for panel B, which is restricted to Brazil). School controls: school location (set of dummies for village, small town, town, city, and large city), student-teacher ratio, log of the number of students, ratio of computers connected to the web as a proxy for school resources, and share of government funding relative to total school funding. Student controls: gender, grade compared to modal grade of students taking the PISA exam in the country, an index of economic, social, and cultural status, and immigration status (set of dummies for native, first generation, and second generation). For control variables, missing variables are replaced with a value of 99 and we include an indicator variable with a value of 1 for each imputed value.

Table 2: Management and student performance, Prova Brasil

| | Portuguese Score | | | | | Mathematics Score | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Management Index | 0.121*** | 0.050*** | 0.036*** | 0.032*** | 0.017*** | 0.134*** | 0.052*** | 0.040*** | 0.036*** | 0.019*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Students | 23829018 | 23829018 | 23829018 | 23829018 | 23829018 | 23827854 | 23827854 | 23827854 | 23827854 | 23827854 |
| Schools | 72683 | 72683 | 72683 | 72683 | 72683 | 72683 | 72683 | 72683 | 72683 | 72683 |
| R-Squared | 0.063 | 0.107 | 0.133 | 0.158 | 0.221 | 0.042 | 0.101 | 0.124 | 0.149 | 0.229 |
| Year FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| State FE | | Y | Y | Y | | | Y | Y | Y | |
| PISA-Like Controls | | | Y | Y | Y | | | Y | Y | Y |
| PB Controls | | | | Y | Y | | | | Y | Y |
| School FE | | | | | Y | | | | | Y |

*Note:* OLS regressions for PB were run with the student-level PB dataset, pooling grades 5 and 9, for years 2007 to 2017. Standard errors clustered at the school level. Test scores are normalized within grade. All specifications include year fixed effects. PISA-like controls are taken from PB data set and attempt to match school controls and student controls in PISA regressions (Table 1): indicator variable for urban schools, student-teacher ratio, log of the number of students, dummies indicating the presence of a computer lab and whether the school has internet access, gender, student households consumption index, and a set of dummies for race. Given the availability of principal characteristics, PB controls include principals age, set of dummies for principals race, principals educational attainment (set of dummies for less than high school, high school, undergraduate (pedagogy), undergraduate (math), undergraduate (Portuguese), undergraduate (others), masters, doctoral), indicator for whether the principal holds another job. PB controls also include the class-year-level share of white teachers, share of teachers holding a college degree, and average teacher tenure. For the students, PB controls include dummies for mother educational attainment (grades 1-5, grades 6-9, secondary grades 10-12, and college). For control variables, missing variables are replaced with a value of 99 and we include an indicator variable with a value of 1 for each imputed value.

Table 3: Management, principal turnover, principal characteristics and student performance, Prova Brasil

| | Portuguese Score | | | | | Math Score | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| **Management** | | | | | | | | | | |
| Management Index (std) | 0.033*** | | | 0.030*** | 0.029*** | 0.037*** | | | 0.033*** | 0.032*** |
| | (0.001) | | | (0.001) | (0.001) | (0.001) | | | (0.001) | (0.001) |
| **Principal characteristics** | | | | | | | | | | |
| New principal | | -0.039*** | | -0.030*** | -0.020*** | | -0.043*** | | -0.032*** | -0.023*** |
| | | (0.002) | | (0.002) | (0.002) | | (0.002) | | (0.002) | (0.002) |
| Gender: male | | | -0.018*** | | -0.015*** | | | -0.012*** | | -0.009*** |
| | | | (0.002) | | (0.002) | | | (0.003) | | (0.003) |
| Age: >50 | | | -0.016*** | | -0.012*** | | | -0.012*** | | -0.008*** |
| | | | (0.002) | | (0.002) | | | (0.002) | | (0.002) |
| Experience: >15yrs | | | -0.045*** | | -0.040*** | | | -0.046*** | | -0.041*** |
| | | | (0.002) | | (0.002) | | | (0.002) | | (0.002) |
| Education: pedagogical degree | | | 0.006*** | | 0.006*** | | | 0.004* | | 0.003 |
| | | | (0.002) | | (0.002) | | | (0.002) | | (0.002) |
| Education: college degree | | | 0.060*** | | 0.051*** | | | 0.060*** | | 0.050*** |
| | | | (0.004) | | (0.004) | | | (0.004) | | (0.004) |
| Wage: >7 min wages | | | 0.034*** | | 0.030*** | | | 0.034*** | | 0.029*** |
| | | | (0.003) | | (0.003) | | | (0.003) | | (0.003) |
| Jobs: only one school | | | 0.022*** | | 0.022*** | | | 0.028*** | | 0.027*** |
| | | | (0.002) | | (0.002) | | | (0.002) | | (0.002) |
| Observations | 22589831 | 22589831 | 22589831 | 22589831 | 22589831 | 22585017 | 22585017 | 22585017 | 22585017 | 22585017 |
| Schools | 72056 | 72056 | 72056 | 72056 | 72056 | 72056 | 72056 | 72056 | 72056 | 72056 |
| R-Squared | 0.188 | 0.188 | 0.188 | 0.188 | 0.189 | 0.160 | 0.159 | 0.160 | 0.160 | 0.161 |
| Year FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| State FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Student Controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Teacher Controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| School Controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

*Note:* OLS regressions for PB were run with the student-level PB dataset, pooling grades 5 and 9, for years 2007 to 2017. Standard errors clustered at the school level. Test scores are normalized within grade. All specifications include year fixed effects. School controls include: whether the school is in an urban area, the student-teacher ratio, school size (log number of students), whether the school has a computer lab an internet access and the level of spending in education in the school's municipality. Teacher controls include: gender, race, education level and tenure at the school. Student controls include: age, gender, socioeconomic status, race, parent literacy and whether the student has a job. Principal characteristics include: gender, age, seniority, whether the principal has a college degree, whether the degree is in pedagogy, and wage band level. Where characteristics are categorical variables in the data, we build indicators that take a value of 1 when categories are above median. For control variables, missing variables are replaced with a value of 99 and we include an indicator variable with a value of 1 for each imputed value.

Table 4: Management and school functioning

| | Teachers | | | | | | Households | |
|---|---|---|---|---|---|---|---|---|
| | (1) shortage | (2) shortage | (3) motivation | (4) motivation | (5) effort | (6) effort | (7) effort | (8) effort |
| **PISA** | | | | | | | | |
| Management Index | -0.056 | -0.057 | 0.299 | 0.316 | 0.048 | 0.071 | 0.257 | 0.282 |
| | (0.023) | (0.023) | (0.027) | (0.025) | (0.024) | (0.023) | (0.029) | (0.027) |
| | [0.016] | [0.013] | [0.000] | [0.000] | [0.046] | [0.002] | [0.000] | [0.000] |
| Observations | 12133 | 12133 | 12133 | 12133 | 12133 | 12133 | 12133 | 12133 |
| Schools | 12133 | 12133 | 12133 | 12133 | 12133 | 12133 | 12133 | 12133 |
| R-Squared | 0.030 | 0.049 | 0.089 | 0.120 | 0.011 | 0.062 | 0.078 | 0.155 |
| Country FE | Y | Y | Y | Y | Y | Y | Y | Y |
| School Controls | | Y | | Y | | Y | | Y |
| Student Controls | | Y | | Y | | Y | | Y |
| **Prova Brasil** | | | | | | | | |
| Management Index | -0.033 | -0.088 | 0.229 | 0.218 | 0.017 | 0.059 | 0.044 | 0.054 |
| | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) | (0.003) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Observations | 322127 | 322127 | 315885 | 315885 | 322273 | 322273 | 322313 | 322313 |
| Schools | 72658 | 72658 | 72321 | 72321 | 72686 | 72686 | 72688 | 72688 |
| R-Squared | 0.001 | 0.448 | 0.052 | 0.377 | 0.000 | 0.490 | 0.002 | 0.481 |
| Year FE | Y | Y | Y | Y | Y | Y | Y | Y |
| PISA-Like Controls | | Y | | Y | | Y | | Y |
| PB Controls | | Y | | Y | | Y | | Y |
| School FE | | Y | | Y | | Y | | Y |

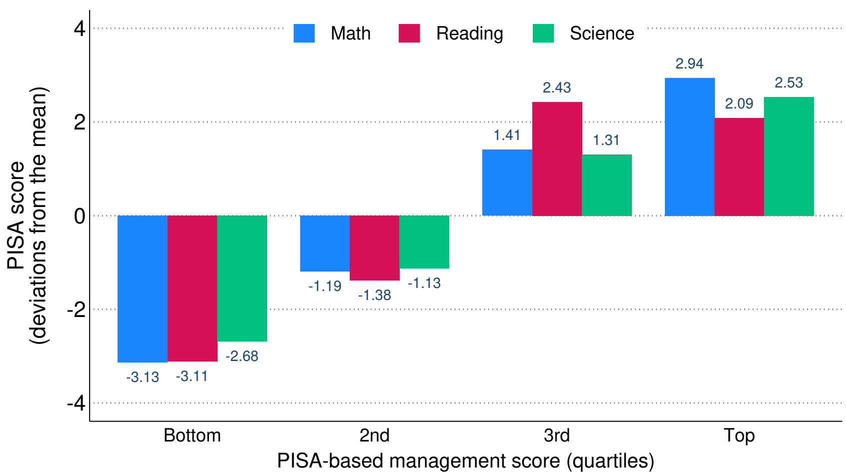
*Note:* Standard errors in parentheses, p-values in square brackets. Panel A: All regressions use data from public schools only. The table reports coefficients from school-level regressions of the PISA-based management index standardized using the overall distribution on each of the intermediate school outcomes (also standardized). All specifications include PISA school final weights and country fixed effects. School controls: school location (set of dummies for village, small town, town, city, and large city), student-teacher ratio, log of the number of students, ratio of computers connected to the web as a proxy for school resources, and share of government funding relative to total school funding. Student controls: gender, grade compared to modal grade of students taking the PISA exam in the country, an index of economic, social, and cultural status, and immigration status (set of dummies for native, first generation, and second generation). Panel B: PB exam is applied in public schools only. The table reports coefficients from school-level regressions of the PB-based management index standardized using the overall distribution on each of the intermediate school outcomes (also standardized). All specifications include year fixed effects. PISA-like controls are taken from PB data set and attempt to match school controls and student controls in PISA regressions (Table 1): indicator variable for urban schools, student-teacher ratio, log of the number of students, dummies indicating the presence of a computer lab and whether the school has internet access, gender, student households' consumption index, and a set of dummies for race. Given the availability of principal characteristics, PB controls include principals' age, set of dummies for principals' race, principals' educational attainment (set of dummies for less than high school, high school, undergraduate (pedagogy), undergraduate (math), undergraduate (Portuguese), undergraduate (others), masters, doctoral), indicator for whether the principal holds another job. PB controls also include the class-year-level share of white teachers, share of teachers holding a college degree, and average teacher tenure. For the students, PB controls include dummies for mother educational attainment (grades 1-5, grades 6-9, secondary grades 10-12, and college). In both panels: For control variables, missing variables are replaced with a value of 99 and we include an indicator variable with a value of 1 for each imputed value.

**ONLINE APPENDIX—NOT FOR PUBLICATION, July 28, 2024**
**"Understanding school management with public data: A new measurement approach and applications, Leaver, Lemos and Scur"**

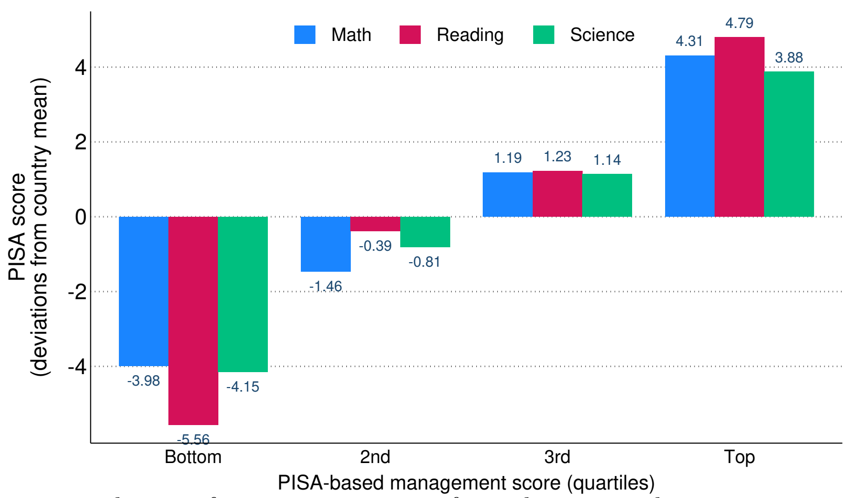
# A Additional Figures and Tables

Figure A.1: Index validation: PISA common questions between 2012 and 2022

(a) PISA 2022, common questions with 2012




(b) PISA 2012, common questions with 2022



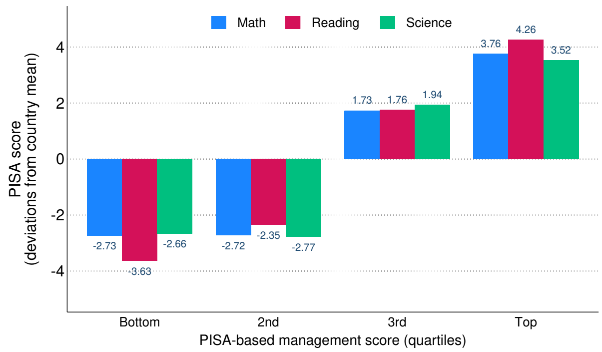

*Note:* This figure uses the set of common questions from the principal questionnaires in PISA 2012 and 2022 that pass the four decision points described in Section 3.1. Student outcomes are estimated using five plausible values and collapsed at the school level using PISA's senate weights. Quartiles of the management index are built at the country level. Test scores are presented as deviations from the subject-specific country mean.

Table A.1: Reproduction of Table 1 from Bloom et al. (2016): Strengths and Weaknesses in Two Ways to Collect Management and Organization Data.

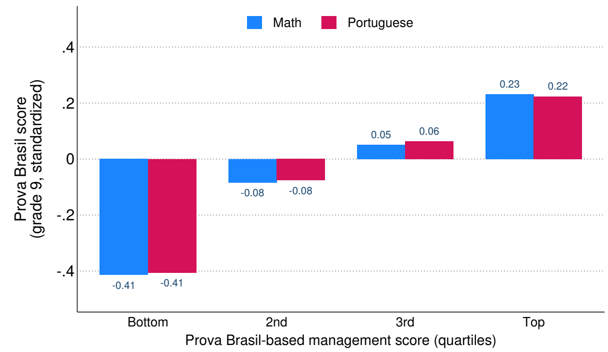
| **Aspect** | **Open (e.g., WMS)** | **Closed (e.g., MOPS)** |
|---|---|---|
| Accuracy of responses | High: Interview format gives opportunity to probe and ask for examples. Possible to implement double-blind method to reduce preconception bias on both sides (interviewer and interviewee). | Medium: Harder to elicit truthful answers if respondents have preconceptions. Greater risk that respondents might misinterpret questions or rush through the survey. |
| Cost per survey | High: High-quality trained interviewers needed to run survey. Training includes one week initial training and ongoing debriefing and calibration. Interviewers time primarily spent recruiting managers to take part in the survey (rather than just running interviews). | Low: Initial design and execution costs, but as this fixed cost can be spread over a very large number of respondents, the cost per survey is low. Costs can be higher in poorer countries where enumerators administer surveys on-site because of unreliable mail and e-mail networks. |
| Response rates | Medium: Interview is interactive and managers more engaged. We obtained an average response rate of 40 percent. | High: Cooperation with a national statistics agency can enable the survey to be mandatory. Given this, response rates are around 80 percent. Without such cooperation, response rates will be low. |
| Replicability | Medium: Training needed to ensure the survey is delivered in same way. Useful to have some individuals who have worked in previous survey waves as trainers for other surveys foster comparability. Training and survey material is available online. | High: Questionnaire essentially the same across countries and already available pre-tested by US Census Bureau. |
| International comparability | High: Multiple countries can be interviewed from same location. Using bilingual interviewers makes it is easier to cross-check responses. | Medium/High: Easier to implement, but there is a risk of differential interpretation if this is not carefully translated across languages. |
| Speed of delivery | High: Can complete a full survey wave in about ten weeks. So, including recruitment and set-up time, possible to complete a survey wave from scratch in about four months. | Medium: Involves cooperation with national statistical agencies, so more planning work in advance. The survey period typically is around three months plus one to three months of data cleaning. |

Figure A.2: Index validation: alternative index building tools

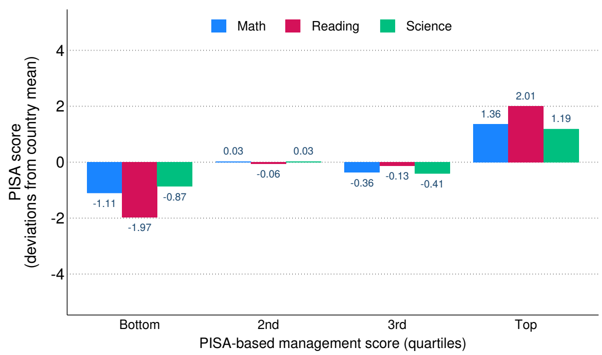
(a) PISA: Anderson (2008) Index



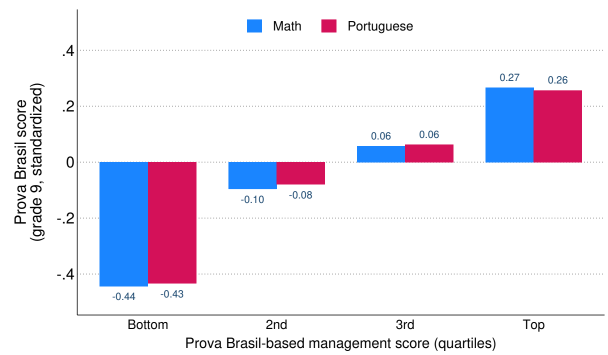

(b) Prova Brasil: Anderson (2008) Index



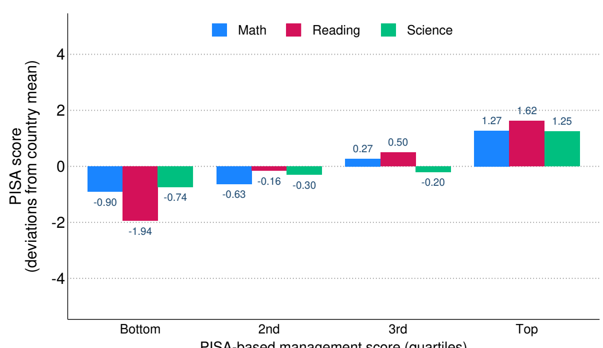

(c) PISA: Principal Component Analysis




(d) Prova Brasil: Principal Component Analysis



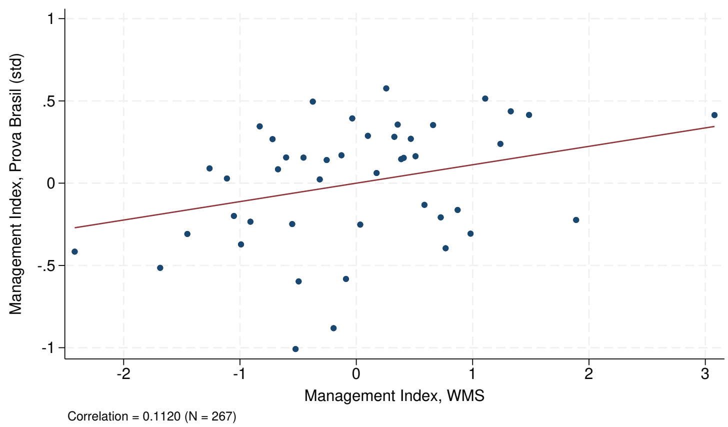

(e) PISA: "Broad" management measure




*Note:* Panels (a), (c) and (e) use PISA 2012 data, 15,196 schools across 65 countries. Student outcomes are estimated using five plausible values and collapsed at the school level using PISA's senate weights. Quartiles of the management index are built at the country level. Test scores are presented as deviations from the subject-specific country mean. Panels (b), (d) and (f) use data from Prova Brasil (2013), grade 9 only. Test scores standardized within subject.

Figure A.3: Index validation, Prova Brasil: school-specific score correlation

(a) Preferred index measure (averages)




(b) Anderson (2008) Index


(c) Principal Component Analysis Index


*Note:* Data from Prova Brasil (2013) and the World Management Survey. The sample contains schools which have data for both Prova Brasil and WMS in 2013, matched at the school level via school identifiers (thus, this sample includes only public schools). Both indices are standardized within-subsample. All graphs are binned scatter plot using 45 quantiles.

Table A.2: Political turnover and school management scores

| | Outcome: Management Index | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Municipal Schools** | | | | | | |
| $1\{IncumbVoteMargin < 0\}$ | -0.211 | -0.202 | -0.229 | -0.222 | -0.262 | -0.249 |
| | (0.040) | (0.039) | (0.056) | (0.054) | (0.044) | (0.043) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Baseline Management Index | 0.344 | 0.308 | 0.337 | 0.298 | 0.333 | 0.297 |
| | (0.011) | (0.011) | (0.014) | (0.014) | (0.012) | (0.012) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Observations | 11306 | 11306 | 6117 | 6117 | 9080 | 9080 |
| R-Squared | 0.150 | 0.169 | 0.149 | 0.170 | 0.144 | 0.162 |
| Clusters | 2454 | 2454 | 1563 | 1563 | 2130 | 2130 |
| Using Bandwidth | 0.139 | 0.139 | 0.070 | 0.070 | 0.110 | 0.110 |
| Optimal Bandwidth | 0.139 | 0.139 | 0.139 | 0.139 | 0.139 | 0.139 |
| **Non-Municipal Schools** | | | | | | |
| $1\{IncumbVoteMargin < 0\}$ | -0.020 | -0.017 | 0.032 | 0.058 | -0.001 | 0.010 |
| | (0.058) | (0.053) | (0.069) | (0.064) | (0.056) | (0.053) |
| | [0.724] | [0.757] | [0.647] | [0.363] | [0.983] | [0.854] |
| Baseline Management Index | 0.405 | 0.373 | 0.384 | 0.349 | 0.391 | 0.359 |
| | (0.017) | (0.015) | (0.018) | (0.017) | (0.015) | (0.014) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Observations | 6714 | 6714 | 3965 | 3965 | 5663 | 5663 |
| R-Squared | 0.178 | 0.198 | 0.166 | 0.189 | 0.172 | 0.193 |
| Clusters | 2118 | 2118 | 1390 | 1390 | 1875 | 1875 |
| Using Bandwidth | 0.133 | 0.133 | 0.070 | 0.070 | 0.110 | 0.110 |
| Optimal Bandwidth | 0.133 | 0.133 | 0.133 | 0.133 | 0.133 | 0.133 |
| Controls | | Y | | Y | | Y |

*Note:* Panel A: Standard errors in parentheses, p-values in square brackets. This table reports the coefficient on political party turnover from regressing standardized management scores in municipal schools on the running variable of the RDD ($IncumbVoteMargin$), political party turnover ($IncumbVoteMargin < 0$), and the interaction of these two variables for the set of municipalities with $IncumbVoteMargin < UsingBandwidth$. We also control for baseline standardized management scores in the year before the election. Controls include school-level controls (whether: the school is located in an urban or rural area, the school is connected to the electric grid, the school is connected to the water network, the school is connected to the sewage system, the schools trash is regularly collected, and the school has Internet) and a 2012 election-cycle indicator. Optimal bandwidth follows Calonico et al. (2017). Panel B: repeat of the analysis in Panel A using non-municipal schools (state and federal schools). Only public schools participate in the Prova Brasil exam. Panel C: repeat of the analysis in Panel A for the municipal schools where the school principal was not replaced. New school principals are those that report being the head of their current school for less than two years on the Prova Brasil school principal questionnaire.

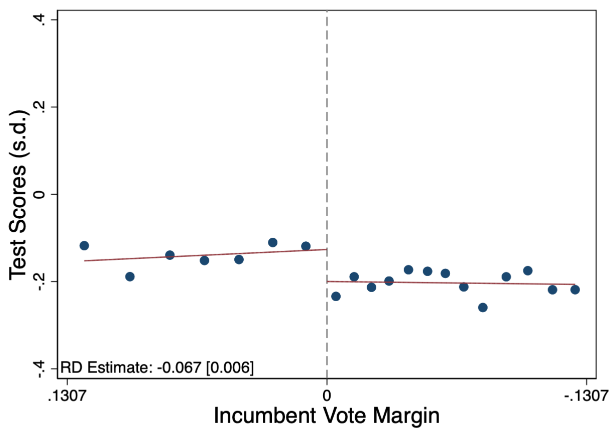
Table A.3: Political turnover and school management scores: by headmaster replacement

| | Outcome: Management Index | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **No Headm. Replacement** | | | | | | |
| $1\{IncumbVoteMargin < 0\}$ | -0.065 | -0.073 | -0.108 | -0.117 | -0.079 | -0.090 |
| | (0.053) | (0.052) | (0.075) | (0.075) | (0.062) | (0.062) |
| | [0.215] | [0.164] | [0.152] | [0.117] | [0.200] | [0.147] |
| Baseline Management Index | 0.373 | 0.353 | 0.354 | 0.332 | 0.362 | 0.340 |
| | (0.016) | (0.016) | (0.021) | (0.021) | (0.019) | (0.019) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Observations | 4337 | 4337 | 2222 | 2222 | 3328 | 3328 |
| R-Squared | 0.194 | 0.204 | 0.185 | 0.197 | 0.187 | 0.198 |
| Clusters | 1376 | 1376 | 804 | 804 | 1143 | 1143 |
| Using Bandwidth | 0.140 | 0.140 | 0.070 | 0.070 | 0.110 | 0.110 |
| Optimal Bandwidth | 0.140 | 0.140 | 0.140 | 0.140 | 0.140 | 0.140 |
| **Headm. Replacement** | | | | | | |
| $1\{IncumbVoteMargin < 0\}$ | -0.284 | -0.260 | -0.227 | -0.212 | -0.285 | -0.263 |
| | (0.056) | (0.054) | (0.068) | (0.066) | (0.055) | (0.053) |
| | [0.000] | [0.000] | [0.001] | [0.001] | [0.000] | [0.000] |
| Baseline Management Index | 0.286 | 0.246 | 0.291 | 0.250 | 0.285 | 0.248 |
| | (0.015) | (0.014) | (0.017) | (0.017) | (0.015) | (0.014) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Observations | 4807 | 4807 | 3387 | 3387 | 5006 | 5006 |
| R-Squared | 0.119 | 0.138 | 0.123 | 0.145 | 0.119 | 0.136 |
| Clusters | 1581 | 1581 | 1187 | 1187 | 1638 | 1638 |
| Using Bandwidth | 0.105 | 0.105 | 0.070 | 0.070 | 0.110 | 0.110 |
| Optimal Bandwidth | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 |
| Controls | | Y | | Y | | Y |

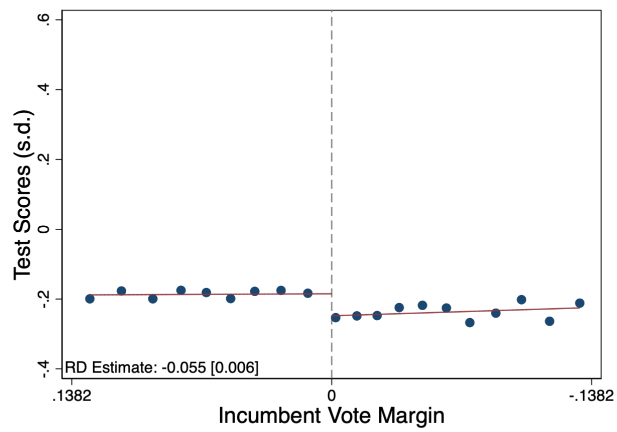
*Note:* Standard errors in parentheses, p-values in square brackets. This table reports the coefficient on political party turnover from regressing standardized management scores in municipal schools on the running variable of the RDD ($IncumbVoteMargin$), political party turnover ($IncumbVoteMargin < 0$), and the interaction of these two variables for the set of municipalities with $IncumbVoteMargin < UsingBandwidth$. We also control for baseline standardized management scores in the year before the election. Panel A restricts the analysis to municipal schools where the principal was not replaced, while Panel B restricts the analysis to municipal schools where the principal was replaced. New headmasters are those that report being the headmaster of their current school for less than two years on the Prova Brasil headmaster questionnaire. Controls include school-level controls (whether: the school is located in an urban or rural area, the school is connected to the electric grid, the school is connected to the water network, the school is connected to the sewage system, the schools trash is regularly collected, and the school has Internet) and a 2012 election-cycle indicator. Optimal bandwidth follows Calonico et al. (2017).

Figure A.4: Political turnover, test scores and school management, by institution

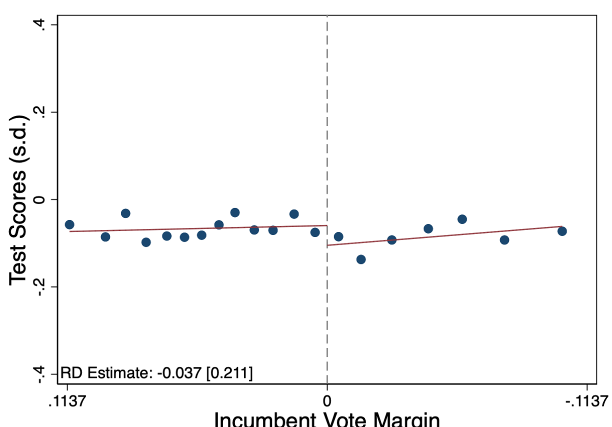
(a) Test scores: principal turnover



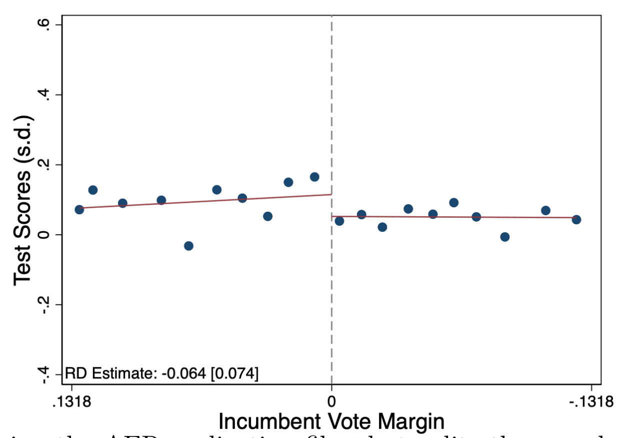

(b) Test scores: appointed structure




(c) Test scores: no principal turnover




(d) Test scores: non-appointed structure




*Note:* This figure replicates Akhtari et al. (2022) using the AER replication files, but splits the sample by whether the municipality's public school staff are politically appointed or selected via a civil service exam. It shows the average of individual-level test scores by bins of $IncumbVoteMargin$ in municipal schools, pooling students from grade 5 and grade 9 and controlling for the average, school-level test scores for the respective grade at baseline. Municipalities with $IncumbVoteMargin < 0$ experienced a change in the political party of the mayor. Municipalities with $IncumbVoteMargin > 0$ did not experience a change in the political party of the mayor. Note that values to the right side of the zero are negative (political turnover), while values on the left side are positive (no political turnover). Selected bandwidth follows Calonico et al. (2017).

# B Appendix: Theoretical Framework

We first present a result that establishes effort levels in high and low management public schools, high management private schools, and the outside sector.

**Lemma 1.** *Assume that the government assigns the teacher to public school $i = L, H$.*

1. *If the teacher accepts the government's offer, then she exerts effort $e^i = \frac{\tau + \Delta^i}{2}$.*
2. *If the teacher declines the government's offer and is hired by a high management private school, then she exerts effort $e^P = \frac{\theta\ B}{2(\bar{\varepsilon} - \underline{\varepsilon})} + \frac{\tau + \Delta^H}{2}$.*
3. *If the teacher declines the government's offer and is hired by an outside employer, then she exerts effort $e^O = \frac{\theta\ \beta}{2(\bar{\varepsilon} - \underline{\varepsilon})}$.*

*Proof.* Part 1. When working in public school $i$, a teacher with baseline motivation $\tau$ chooses effort to solve

$$\max_e\ G - (e^2 - (\tau + \Delta^i) \cdot e).$$

Differentiation to obtain the first order condition yields the solution stated above. (Here, as in the cases below, the second order condition necessary for a maximum holds.)

Part 2. When working in a high management private school, a teacher with baseline motivation $\tau$ and ability $\theta$ chooses effort to solve

$$\max_e\ P \cdot B + W - (e^2 - (\tau + \Delta^H) \cdot e)$$

where $P$ is the probability that $y_1^H$ exceeds the threshold $\bar{y}$ given teacher and household effort. Using the uniform distribution for $\varepsilon$, we can rewrite this probability as

$$P = \Pr\left(\theta\ e + a + \varepsilon > \bar{y}\right) = \Pr\left(\theta\ e + a - \bar{y} > -\varepsilon\right) = \frac{\bar{\varepsilon} + \theta\ e + a - \bar{\bar{y}}}{\bar{\varepsilon} - \underline{\varepsilon}}.$$

The first order condition for this optimization problem is

$$\frac{\theta\ B}{\bar{\varepsilon} - \underline{\varepsilon}} = 2e - (\tau + \Delta^H),$$

which yields the solution stated above.

Part 3. When working in the outside sector, a teacher chooses effort to solve

$$\max_e\ P^O \cdot \beta - e^2,$$

where $P^O$ is the probability that $z$ exceeds the threshold $\bar{z}$ given effort. We can rewrite this probability as

$$P^O = \Pr\left(\theta\ e + \varepsilon^O > \bar{z}\right) = \Pr\left(\theta\ e - \bar{z} > -\varepsilon^O\right) = \frac{\bar{\varepsilon} + \theta\ e - \bar{z}}{\bar{\varepsilon} - \underline{\varepsilon}}.$$

The first order condition for this optimization problem is

$$\frac{\theta\ \beta}{\bar{\varepsilon} - \underline{\varepsilon}} = 2e,$$

which yields the solution stated above. □

We now use these effort levels to construct Figure B.1. Calculations were performed in Mathematica; the notebook file is available on request.

**Derivation of Figure B.1** The figure is based on the following numerical example. Teacher intrinsic motivation is distributed $\tau \sim U\,[0, 10]$, and teacher ability is distributed $\theta \sim U\,[1, 5]$. In the low management public school $G^L = 30$, $\Delta^L = 0$, and $\gamma^L = 1$. In the high management public school, $G^H = 35$, $\Delta^H = 0.5$ and $\gamma^H = 2$. Other parameters are set at $W = 15$, $B = 40$, $\bar{y} = 4.5$, $\beta = 50$, and $\bar{z} = 1$.

The unshaded region in the top panel of Figure 7 shows $\mathcal{T}^H$, the set of $(\theta, \tau)$ types for whom the payoff from accepting a job in the assigned high management public school (weakly) exceeds both the expected payoff of declining and accepting a job in a high management private school and the expected payoff of declining and accepting a job in the outside sector. This region is bounded by two functions

$$\tau_P^H = \frac{7}{\theta} - 2\theta - \tfrac{1}{2}, \quad \tau_O^H = \sqrt{25\theta^2 - 60} - \tfrac{1}{2}.$$

The function $\tau_P^H$ traces out the loci of $(\theta, \tau)$ types who, anticipating subsequent teacher and household effort, are indifferent between accepting the job in the assigned high management public school and declining it in favour of a job in a high management private school, i.e. types for whom

$$G - (e^H)^2 + \left(\tau + \Delta^H\right)\, e^H = W + B\left(\frac{\overline{\varepsilon} + \theta\, e^P + a^P - \overline{y}}{\overline{\varepsilon} - \underline{\varepsilon}}\right) - (e^P)^2 + \left(\tau + \Delta^H\right)\, e^P.$$

Substituting for $e^H$ and $e^P$ from Lemma 1, together with the parameters in the numerical example (implying $a^P = 1$), and rearranging yields the expression for $\tau_P^H$ stated above. Fixing $\theta$, for any $\tau < \tau_P^H(\theta)$, the teacher's payoff from accepting the government's offer is strictly higher than her expected payoff from declining and accepting a job in a high management private school.

The function $\tau_O^H$ traces out the loci of $(\theta, \tau)$ types who, anticipating subsequent teacher effort, are indifferent between accepting the job in the assigned high management public school and declining it in favour of a job in the outside sector, i.e. types for whom

$$G - (e^H)^2 + \left(\tau + \Delta^H\right)\, e^H = \beta\left(\frac{\overline{\varepsilon} + \theta\, e^O - \overline{z}}{\overline{\varepsilon} - \underline{\varepsilon}}\right) - (e^O)^2.$$

Substituting for $e^H$ and $e^O$ from Lemma 1, together with the parameters in the numerical example, and rearranging for $\tau$ yields the expression for $\tau_O^H$ stated above. Fixing $\theta$, for any $\tau > \tau_O^H(\theta)$, the teacher's payoff from accepting the government's offer is strictly higher than her expected payoff from declining and accepting a job in the outside sector.

The values for average ability and average baseline intrinsic motivation (the coordinates of the blue dot) are obtained by numerical integration.

The unshaded region in the bottom panel of Figure 7 shows $\mathcal{T}^L$, the set of $(\theta, \tau)$ types for whom the payoff from accepting a job in the assigned low management public school (weakly) exceeds both the expected payoff of declining and accepting a job in a high management private school and the expected payoff of declining and accepting a job in the outside sector. This region is bounded by two functions

$$\tau_P^L = \frac{36}{8\theta+1} - 2\theta - \tfrac{1}{4}, \quad \tau_O^L = \sqrt{25\theta^2 - 40}.$$

The function $\tau_P^L$ traces out the loci of $(\theta, \tau)$ types who, anticipating subsequent teacher and household effort, are indifferent between accepting the job in the assigned low management public school and declining it in favour of a job in a high management private school, i.e. types for whom

$$G - (e^L)^2 + \left(\tau + \Delta^L\right) e^L = W + B\left(\frac{\overline{\varepsilon} + \theta\, e^P + a^P - \overline{y}}{\overline{\varepsilon} - \underline{\varepsilon}}\right) - (e^P)^2 + (\tau + \Delta^H)\, e^P.$$

Substituting for $e^L$ and $e^P$ from Lemma 1, together with the parameters in the numerical example (implying $a^P = 1$), and rearranging yields the expression for $\tau_P^L$ stated above. Fixing $\theta$, for any $\tau < \tau_P^L(\theta)$, the teacher's payoff from accepting the government's offer is strictly higher than her expected payoff from declining and accepting a job in a high management private school.

The function $\tau_O^L$ traces out the loci of $(\theta, \tau)$ types who, anticipating subsequent teacher effort, are indifferent between accepting the job in the assigned low management public school and declining it in favour of a job in the outside sector, i.e. types for whom

$$G - (e^L)^2 + (\tau + \Delta^L)\, e^L = \beta\left(\frac{\overline{\varepsilon} + \theta\, e^O - \overline{z}}{\overline{\varepsilon} - \underline{\varepsilon}}\right) - (e^O)^2.$$

Substituting for $e^L$ and $e^O$ from Lemma 1, together with the parameters in the numerical example, and rearranging for $\tau$ yields the expression for $\tau_O^L$ stated above. Fixing $\theta$, for any $\tau > \tau_O^L(\theta)$, the teacher's payoff from accepting the government's offer is strictly higher than her expected payoff from declining and accepting a job in the outside sector.

The values for average ability and average baseline intrinsic motivation (the coordinates of the orange dot) are also obtained by numerical integration.

**Low-cost private schools** We complete the analysis by considering an alternative numerical example, where pay in a high management private school is *below* the level in both high and low management public schools. All parameters take the same values as in the previous numerical example, except $W = 5$ and $B = 20$. In this numerical example,

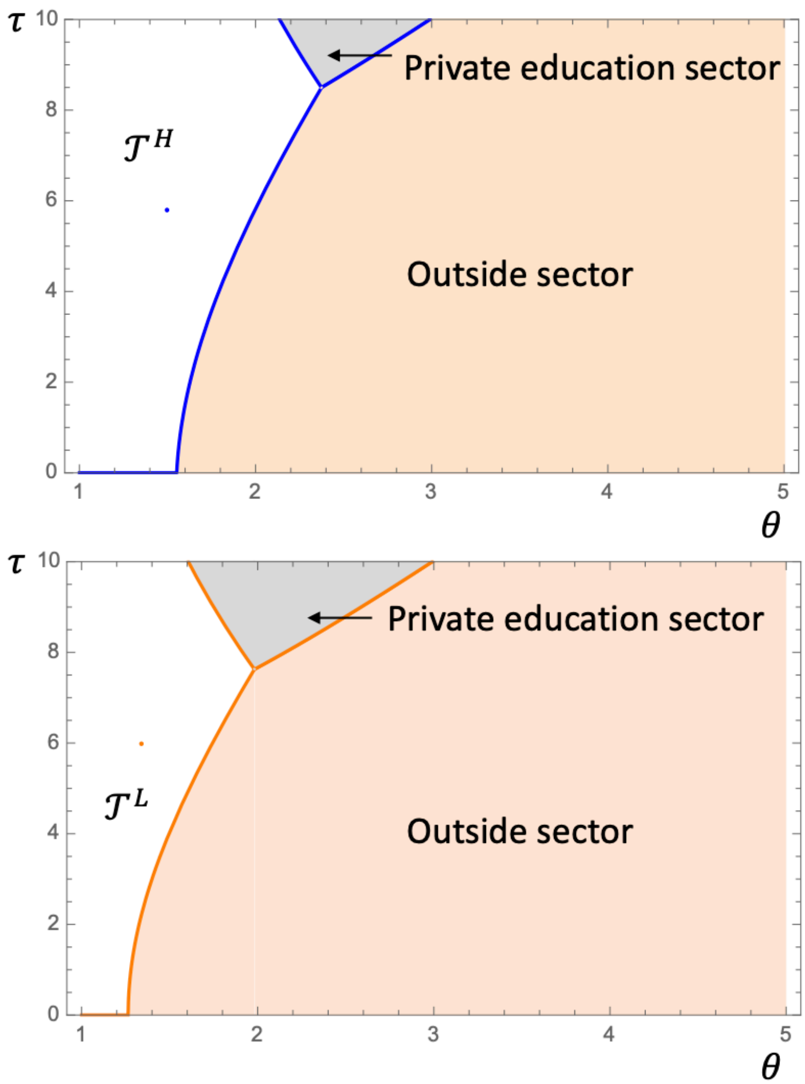
$$\tau_P^H = \frac{27}{\theta} - \theta - \tfrac{1}{2}, \quad \tau_O^H = \sqrt{25\theta^2 - 60} - \tfrac{1}{2}$$

and

$$\tau_P^L = \frac{88}{4\theta+1} - \theta - \tfrac{1}{4}, \quad \tau_O^L = \sqrt{25\theta^2 - 40}.$$

These functions are plotted in Figure 9. As before, the probability of hiring the teacher in a high management public school is higher than the probability of hiring the teacher in a low management public school (the unshaded region is larger in the top panel than in the bottom panel). The expected intrinsic motivation of a teacher hired to a high management public school is now slightly *lower* than the expected intrinsic motivation of a teacher hired to a low management public school (compare the height of the orange dot at 5.98 with the height of the blue dot at 5.81). The difference is small, however, and not sufficient to reverse the effort effect: the expected effort level of a teacher hired to a high management public school is higher than the expected intrinsic motivation of a teacher hired to a low management public school ($\mathrm{E}\left[\frac{\tau+\Delta^H}{2}|(\theta,\tau)\in\mathcal{T}^H\right]=3.16>\mathrm{E}\left[\frac{\tau+\Delta^L}{2}|(\theta,\tau)\in\mathcal{T}^L\right]=2.99$). Household effort levels in public schools are unchanged.

Figure B.1: Teacher selection, with 'low cost' private schools




*Note:* Teacher ability is distributed $\theta \sim U[1,5]$ and teacher intrinsic motivation is distributed $\tau \sim U[0,10]$. In the low management public school $G^L=30$, $\Delta^L=0$, and $\gamma^L=1$. In the high management public school, $G^H=35$, $\Delta^H=0.5$ and $\gamma^H=2$. Other parameters are set at $W=5$, $B=20$, $\bar{y}=4.5$, $\beta=50$, and $\bar{z}=1$. The blue point in the top panel shows average teacher ability and average baseline intrinsic motivation among teacher types who select into a high management public school. The $(x,y)$-coordinates are $(1.50, 5.81)$. The orange point in the bottom panel shows average teacher ability and average baseline intrinsic motivation among teacher types who select into a low management public school. The $(x,y)$-coordinates are $(1.34, 5.98)$.

# C Variable classification: PISA and Prova Brasil

## C.1 PISA

## C.2 Prova Brasil