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

Clare Leaver* University of Oxford

Renata Lemos† World Bank

Daniela Scur‡ Cornell University

This draft:
December 2025

Abstract

Why do students learn more in some schools than others? One consideration receiving growing attention is school management. This paper presents a theoretical framework to explain how school management could contribute to student learning. In parallel, we develop a new approach to measuring management practices at scale using existing public data and exemplify the methodology with OECD's PISA. Two validation exercises confirm a strong, positive relationship between our new management index and student learning, and support for the predictions of the theoretical model. We apply our measurement approach to unpack the negative relationship that has been documented between managerial turnover and organizational performance. We provide new evidence from Brazilian schools that shows turnover matters, at least in part, due to a de-stabilizing effect that reduces the quality of management practices which, in turn, lowers student test scores.

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, Rafael Rocha-Lima Bentes and Felipe Macedo Dias for excellent research assistance. Leaver is grateful for the hospitality of Toulouse School of Economics, 2018-19.

*Blavatnik School of Government. CEPR and J-PAL. 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. SC Johnson College of Business. Charles H. 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 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 why school management leads to better student learning; and (b) how to measure school management accurately and cost-effectively at scale and across contexts.

This paper addresses these 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 and draw out testable predictions that can be taken to international data.

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 the OECD's Programme for International Student Assessment (PISA). We start with a decision tree and use it to identify management practices in the PISA school principal survey. We then show how the resulting 49 questions can be coded and built into a school management index. Our PISA-based index covers over 15,000 schools across 65 countries. We validate this index by documenting that it is strongly positively correlated with student performance, echoing findings based on the well-established World Management Survey (Bloom et al., 2015). As a further validation exercise, we use the PISA dataset to construct indices of school functioning and show that they are correlated with our PISA-based management index, as predicted by the theoretical model.

We hope there will be myriad uses of this new measurement approach. To demonstrate its potential value, we apply the method to a second dataset, Brazil's *Prova Brasil*, and use the resulting management index to unpack the negative relationship that has been documented between leadership change and student performance (Akhtari et al., 2022).[1] We explore this

[1] Leadership change is notoriously disruptive. Though the wider literature on the impact of leadership change often focuses on CEOs (Bandiera et al., 2020; Bertrand and Schoar, 2003; Huber et al., 2021), causal evidence is scarce, typically relying on rare events such as CEO deaths or wars. Our analysis exploits a commonplace event, political turnover, that brings with it leadership change in schools.

issue in two stages. First, we test whether an exogenous turnover shock impacts school management practices. Building on the regression discontinuity design in Akhtari et al. (2022), we show that upheaval in school leadership due to changes in municipal governments following close elections led to a subsequent decline in the quality of school management. Next, we use mediation analysis to quantify the relative importance of management practices and principal characteristics (attributes such as gender, age, experience, and educational background that may shape leadership prowess or charisma) as causal pathways from turnover to test scores. Following the approach of Acharya et al. (2016), we find that management practices and principal characteristics together account for close to a quarter of the original negative partial effect of turnover on test scores, with management playing the larger mediating role. We conclude that turnover matters, at least in part, through a de-stabilizing effect that reduces the quality of school management practices which, in turn, lowers student test scores.

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., 2024; 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 application, facilitates quasi-experimental methods for evaluating, e.g., the impact of turnover in political and organizational leadership. A further benefit is that researchers running experiments now have a larger set of benchmarkable questions on management practices to consider for inclusion in baseline and endline surveys (c.f. 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 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

[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).

tracking (Hanushek and Wössmann, 2006; Ruhose and Schwerdt, 2016), external exams (Wössmann, 2005), and instructional time (Lavy, 2015). Our new indices 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 models of Biasi (2021) and Brown and Andrabi (2025). 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. Our framework can be (and indeed has been) used by researchers with access to the data necessary to explore the causal impact of management practices in schools.[3]

2 Why might management matter in schools?

It is now well established that more structured 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, noting testable predictions that can then be taken to the data.[4]

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. In other contexts, there has been a growth of 'low-cost' private schools and jobs in the public sector confer significant rents. 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.

The framework focuses on a teacher and representative household. 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. 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.

[3]See, e.g., Prem and Munoz (2023) who use the framework in a study of principal recruitment in Chile.

[4]We summarize the model and give intuition for results. Details and derivations are in the Appendix.

2.1 Summary of the model

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 and effort. When working in the education sector, she (initially) derives a marginal benefit from exerting an extra unit of effort, e, in teaching. This intrinsic motivation has two components: a baseline level, τ, that is private information; and a motivational increment, Δ, that is public information and determined by the management practices in the teacher's chosen school. Intrinsic motivation plays no role in the outside sector. Our model also includes a representative household (student plus parents) that cares about its effort, a. The household's intrinsic motivation depends on γ, an increment that is determined by school management practices.

We use y_1 to denote a representative student's learning outcome in a school that hires the teacher, and y_0 to denote the same in a school that does not hire the teacher. We assume $y_1 = \theta e + a + \varepsilon$ and $y_0 = a + \varepsilon$. If the teacher chooses to work in the outside sector, her performance is $z = \theta e + \varepsilon$. The term θ is the teacher's ability and is her private information, and the term ε is an error that is independent across employments. Public schools offer a flat wage. Private schools and the outside sector offer a base wage plus a performance bonus.

We assume that management has three effects. Good management practices: (i) enable managers to cultivate the intrinsic motivation of their staff, increasing Δ; (ii) free up resources and enable managers to offer a higher level of pay (potentially in hedonic form); and (iii) help to create a stimulating environment for students and parents, increasing γ. 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, Δ, base pay, and γ are all higher than in a low management school.

2.2 Mechanisms

We use this model 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 (θ, τ) 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 $\mathrm{E}[y^i]$ to denote the expected learning outcome (*ex ante*, prior to occupational and effort choices) in school i, and $\mathcal{T}^i$ to denote the set of (θ, τ) types that can be hired to this school. In the Appendix, we derive teacher and household effort in high and low management public schools. Substituting for these expressions, we can decompose the difference in expected learning outcomes between high and low management public schools as

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

where $1_{\{(\theta,\tau)\in\mathcal{T}^i\}}$ is an indicator function for the event that the teacher is hired to school i.

The first term on the RHS of equation (1) is what we call the *teacher incentive effect* of good management. Here, we compare the expected teacher contribution to the learning 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 (θ, τ) types fixed and just consider how the incentive environment for the teacher contributes to learning. This expression shows that expected learning is higher in a high management public school, in part, because good management practices increase the intrinsic motivation of any given (θ, τ) 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 call the *teacher selection effect* of good management practices. Here, we compare the expected teacher contribution to the learning 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. Expected learning is higher in a high management public school, in part, because good management practices encourage better (θ, τ) 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.[5]

[5]Figure B.1 in the Appendix provides an illustration of this teacher selection effect. The unshaded area in the top panel depicts the set of (θ, τ) types that are hired by a high management public school, while the unshaded area in the bottom panel depicts the set of (θ, τ) 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

The third term in equation (1) captures what we call the *household incentive effect* of good management practices. Expected learning 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.

These mechanisms — teacher incentives, teacher selection and household incentives — could explain the widely documented positive correlation between management scores and student learning outcomes. If these mechanisms are correct, then we should see behavioral responses in school functioning, as set out in the four predictions below.

Prediction 1: Teacher shortages. The theory 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 theory 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).

Prediction 3: Teacher effort. The theory 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 theory 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).

In Section 3, we use this theoretical framework to ground the development of indices of school functioning based on public data. We will show that all these predictions have support in the data we build, suggesting that the causal pathways from the quality of management practices to student learning posited in the theory are empirically plausible across a wide range of countries. While our evidence is only correlational, it provides validation for the framework, which can be used to explore the *causal* impact of management practices on school outcomes in other settings (see, e.g., Prem and Munoz, 2023).

---

school: the unshaded area is larger in the top panel relative to the bottom panel. It also hires *better* types: average θ and average τ, shown by the (x, y) coordinates of the blue dot, are higher relative to the bottom panel.

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 (see Scur et al. (2021) for a recent survey). In education research, elements of management were imbued in a black box of school or principal "fixed effects", often bundling manager and management factors.[6] Over the last two decades, there have been substantial advances in measurement of management practices but challenges still remain. Below, we summarize existing methods and then introduce our new approach.

3.1 Modes of data collection

Phone surveys: accurate but not scalable. The leading example is the World Management Survey (WMS): a large, cross-country project that measures the adoption of management best practices in multiple sectors using a detailed, but expensive, survey methodology. It was originally developed to measure the adoption of structured management practices in manufacturing firms (Bloom and Van Reenen, 2007) and was subsequently adapted to other areas, including the public sector (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 is closely monitored by supervisors.

The WMS school survey features 20 topics, covering instructional processes, use of data in assessments, target-setting and personnel practices. 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). 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.[7]

[6] 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."

[7] The version for developing countries (D-WMS, from Lemos et al., 2024) has been deployed in Brazil, Colombia, Haiti, Indonesia, Mexico, Mozambique, Ghana, Pakistan, Puerto Rico, Tanzania and Vietnam. The core result showing a positive relationship between management and student performance is seen across countries.

One of the advantages of the WMS methodology is that an independent analyst who interviews dozens of principals (including across countries and languages) scores the practices for each school they interview, as well as a share of schools other analysts interview. This consistency allows for cross-country comparability, but it comes at a high cost (approximately USD400 per interview) and can be slow (about 4 months to conduct a single country wave). The weaknesses are that it might not suit every context and it is inherently not scalable (Scur et al., 2021).

Self-respondent surveys: scalable but noisier. Partially in response to this lack of scalability, a sister project was conceived at the US Census Bureau. 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 (MOPS) that is now sent to over 35,000 manufacturing firms in the US (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 (Scur et al., 2024). But, scale is necessary to account for the additional noise stemming from self-responses. A further issue is that it is not possible to objectively account for cultural biases in responses.

Our approach. 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 (see Section 3.5), 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, contributing to accessibility and reproducibility efforts. It can also provide new projects with ready-to-go sets of questions to measure management and existing, benchmark data. 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 a popular dataset in education research: OECD's PISA.

3.2 A decision tree for identifying management practices in existing surveys

We focus primarily on the school principal questionnaire from PISA. We review each survey question and take it to the four steps in our decision tree, whether the question: (1) is objective; (2) asks about a process; (3) is linked to a management practice and (4) has a WMS benchmarking equivalent. We illustrate the thought process using a question from PISA 2012:

> 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). 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, or descriptions of allocation of authority and principal autonomy are not classified as 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 (i.e. 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, most questions we identified as being objective and describing a process were also linked to a management practice. The exception were the set of questions describing processes of decision-making authority and school autonomy. Below is an example of a question that we classified as a process that is not a management practice (Example 4); the question is not asking *how* selection of teachers is done, but rather *who* has responsibility for each process. This set of questions could be its own index of autonomy, but we do not include them as part of the core management index.[8]

> Example 4: *"Regarding your school, who has a considerable responsibility for the following tasks?"*
>
> (a) *Selecting teachers for hire.*
>
> – *PISA 2012, Question 29(a) from the Principal Questionnaire*

Node 4: Does the topic addressed in the survey question have a WMS equivalent? The final decision point (for our preferred index in this paper) is whether the topic addressed by the question has a WMS equivalent. We include this node because, as this is a new methodology, having a strong benchmark to ground the exploration is important. 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 if there are reasons to expect the monotonicity to hold.

3.3 Building and validating a management index using PISA

We run this exercise using PISA 2012 (OECD, 2021). In total, PISA 2012 has 55 questions that pass the first three decision points, and 49 of them have counterpart topics in the WMS

[8]Indeed, there are autonomy questions not unlike these in the WMS and they are not part of the main management index. More generally, it is not clear *ex ante* that more autonomy is always better; evidence suggests that it depends on the context and the level of state capacity that schools operate in (e.g. Fuchs and Wössmann, 2007; Jackson, 2025).

(and hence pass the fourth decision node).[9] For each question, we code the responses to a value between 0 and 1, as in the MOPS Census questionnaire, with responses closer to 0 indicating 'less structure' — that is, issues are dealt with 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 5: *"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.*
>
> 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.*
>
> – *PISA 2012, Question 30(t) from the Principal Questionnaire*

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.

We conduct this coding exercise with every question identified as relevant via the decision tree. We then average the scores within each of the WMS-equivalent practices. Finally, we build our management index using the same approach as the WMS, for each country: standardize each practice score, take the average across the topics, and then standardize again. The reason we first aggregate questions into the WMS-equivalent practice is to avoid over-weighting any one particular topic for which there are multiple questions measuring the

[9]This also coincides with the 'best' year for management questions in the PISA 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 (see Liberto et al., 2015).

same construct in a survey. For example, if there are five questions about a performance tracking practice our method will average them first, such that they will be weighted the same as other practices that might have fewer questions in the survey. Robustness checks, where we vary this and other elements of the index-building methodology are presented in Section 3.4. First, we present two validation exercises where, in the spirit of Bloom et al. (2015), we ask if our index is measuring anything of substance relative to outcomes that we care about: student learning and (per our theoretical framework) school functioning.

Correlation with student performance. We start with a simple check of the raw data. Figure 1 plots the average test scores of students, in deviations from the country-mean, for each quartile of the new PISA management index. We find that students in schools in the bottom quartile of (within-country) management score are, on average, performing about 1.5 to 2.5 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 2.5 points higher than their country's mean. To put this into context, PISA is standardized across years and countries such that the mean is 500 and the standard deviation is 100, and 40 points on the PISA scale is equivalent to about one year of learning (OECD, 2019). The range of these results mirrors 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).

To check the conditional correlation between our PISA management index and student performance, we estimate the following regression model:

$$Y_{isc} = \beta_0 + \beta_1 \; MGMT_{sc} + SCH^{'}_{sc} \; \beta_2 + STU^{'}_{isc} \; \beta_3 + \gamma_c + \varepsilon_{isc}, \tag{2}$$

where Y_{isc} is a test score outcome (reading, math or science) for student i in school s and country c in the PISA 2012 assessment. $MGMT_{sc}$ is the management index for school s in country c, and γ_c is a country fixed effect. The row vectors SCH_{sc} and STU_{isc} contain school-level and student-level controls (full list of variables reported in table notes). Standard errors are clustered at the school level. To estimate this equation we use the OECD's `repest` Stata command, which relies on the five test score 'plausible values' for each student and subject. The estimator uses OLS and a series of survey weights to obtain standard errors that account for both sampling and measurement error.[10] The parameter of interest in this

[10]See Jerrim et al. (2017) for a thorough review of how to best use PISA scores and survey weights. Given the complex methodology of the PISA assessment and stratification choices, it is important to use the `repest` command to avoid incorrectly estimating standard errors.

exercise is β_1.

Table 1 reports the results. Column (1) shows the relationship between our management index and reading test scores, controlling only for country fixed effects. The coefficient indicates that one standard deviation higher score on our management index is associated with an increase in reading performance equivalent to about four-weeks-worth of learning. Column (2) includes school controls, which absorb some of the variation, and suggest a correlation that is equivalent to about 1.2 weeks of learning. Column (3) includes student controls, which further reduces the coefficient, albeit by just half a week of learning. Columns (4) to (9) repeat the exercise for math and science test scores and the pattern is similar. While only correlational, these results suggest that our management index is measuring something meaningful about the processes at these schools, echoing the findings in Bloom et al. (2015).

Correlation with school functioning. Our theoretical framework proposes mechanisms through which school management practices could shape student outcomes. We noted above that, if these mechanisms are correct, we should see evidence in how schools function, specifically in teacher shortages, motivation and effort, as well as household effort. As a further validation exercise, we look to the correlation of our PISA management index with measures of these intermediate outcomes.

We exploit another advantage of a large public dataset such as PISA, namely that it contains additional data relating to teachers, students and parents. We construct four measures of school functioning using questions from the PISA principal survey (details on this index construction are in the Appendix). We do not build a decision tree as we did for the management index because (a) we are open to including both subjective and objective questions, and (b) we do not have strong priors as to which questions are more or less important. Instead, we simply include all questions that relate to each of the constructs that we are seeking to measure. We standardize these indices within-country. We check the conditional correlation between our PISA management index and each of these four indices using the school-level analogue of the regression model in equation (2).

Table 2 reports the results. The dependent variable in the first two columns is the teacher shortage index. Column (1) includes only country fixed effects. Column (2) adds school and school-level averages of 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 0.058 standard deviation decrease in the teacher shortage index. The remaining columns repeat this exercise for the teacher motivation, teacher effort, and household effort indices. In all columns, the relationship between our PISA management index and our PISA index of school functioning is significant. All partial correlations are positive, as predicted by the theoretical model. We view these results as further evidence that our management index is able to capture meaningful variation in processes in schools. These findings also lend support to the model and mechanisms outlined in Section 2.

3.4 Robustness

Although our methodology is grounded in prior literature and benchmarking (e.g. Bloom et al., 2015; Lemos et al., 2024), the process of identifying which questions in a given survey should qualify for a management index is still fundamentally researcher-driven. To understand how these choices affect the final index, we run a series of robustness checks on our decision tree and the build process itself.

3.4.1 Impact of decision nodes

We first iterate through each decision node and drop the restriction, re-build the index including the previously excluded questions from that node, and re-run the first validation exercise based on equation (2) above. Although there is a strong rationale for our preferred approach (use only questions that satisfy all four decision nodes), it is important to know if and how deviations matter.

Table 3 reports the results. Column (1) repeats the base specification (including all controls) as a comparison. Column (2) includes questions restricted by Decision Node 4 (WMS equivalent) to the preferred index. Column (3) includes questions restricted by Decision Node 3 (linked to a management practice). In PISA 2012, this node only excludes questions that measure principal autonomy across a range of processes.[11] Column (4) includes questions restricted by Decision Node 2 (is it a process). In PISA 2012, this primarily excludes questions that describe school policies, without being explicit about implementation. For all these variants, adding each set of variables to the preferred index has a similar impact: a slight reduction in the coefficient and the correlation remains significant.

In sum, including a larger number of more ambiguous constructs attenuates the core relationship between management and test scores. Our recommendation is therefore to follow

[11] This is the case in PISA 2012, but other surveys may include types of questions that are objective and are processes (nodes 1 and 2) but not linked to a management practice (say, pedagogical practices, as noted in the discussion above). As such, the node should not be interpreted as only referring to the exclusion of autonomy-related questions.

the decision tree proposed in this paper. Researchers may wish to explore the relationship between these additional constructs and student outcomes in future work.

3.4.2 Alternative index-building methods

Since our goal was to benchmark against the existing 'state of the art', we followed the spirit of the WMS index-building method. For robustness, we verify whether our index and core relationships are robust to three alternative index-building methods: averaging across all questions without bundling into practices first; aggregating into the first latent factor using Principal Component Analysis; and constructing the Anderson (2008) index. Results are reported in Columns (5) to (7) in Table 3. The general patterns are similar using these different index-building methods, suggesting the underlying construct we are measuring is picked up by indices independently of how they are built.

3.4.3 Individual topics: leave-one-out exercise

Starting from our preferred index building method, we probe whether one question or set of questions could be the primary driver of the variation in the index. Here, we exclude one practice at a time, re-build the index and run the main validation student outcome regressions (as per the specifications in Columns 3, 6 and 9 of Table 1). We find that 'leaving-one-out' changes the coefficient and significance level slightly, though it is mostly a minor impact. This suggests the core relationship is not driven by any one single practice. The vast majority of coefficients remained within a 5 percent significance level, and all remained within the 10 percent significance level. Figure A.2 in the Appendix reports the distribution of the estimated coefficients on the management index for each of the indices with one question 'left out'.

3.4.4 Data-driven index-building: exploratory factor analysis

An alternative to choosing questions via our decision tree is to take a data-driven approach such as exploratory factor analysis — a method commonly used to identify 'latent constructs' (i.e. factors) that explain the patterns of correlations among a set of variables, without imposing any pre-specified structure. We input variables constructed from all PISA questions that were not characteristics (i.e. number of students, location, etc) or intermediate outcomes (teacher motivation, etc) into the model. For PISA 2012, the set of variables thus includes all variables we had identified as management, as well as the autonomy and policy variables. The model returns six factors across all countries that capture over 70 percent of the shared variance. We then predict six factors within each country dataset and correlate each factor with student test scores.[12]

Table 4 reports the results.[13] The first and second latent factors are significantly correlated with test scores, though one is positively and one is negatively correlated. The remaining factors are not significantly correlated with outcomes. Labeling these factors is not straightforward: the first likely reflects a latent factor of 'general good management', similar to the factor analysis results using the World Management Survey data, as the majority of variables load positively on this factor (c.f. Scur et al., 2021). The interpretation for the second factor is less clear, though it seems to load mostly negatively on the management variables, mixed on autonomy and mostly positively on policy variables. Beyond these two factors, variables load positively and negatively without following a discernible pattern. Indeed, when we correlate each of the first six factor indices with our preferred index there is a positive yet noisy correlation between estimated Factor 1 and our preferred index, but no correlation whatsoever with any of the other factors (Figure A.3).

This exercise highlights the benefit of our literature-driven mode of variable selection. Focusing directly on the variables that should (based on prior work) matter yields a stronger and more precise metric. Our decision tree is also more transparent and enables users to understand what goes in and (hence) interpret what comes out.

3.4.5 Extension to other datasets: Prova Brasil example

Our methodology can be used to build management indices based on numerous other surveys. Many countries conduct national school surveys alongside administration of standardized tests. Latin America, for example, is particularly prolific: 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.[14] To illustrate, we apply our methodology to Prova Brasil.[15] The patterns of the core relationship between management and student performance are strikingly similar to those we found using PISA data. We also find a positive correlation

[12]Five countries (Argentina, Germany, Hong Kong, New Zealand and Singapore) had a large number of missing values in the autonomy questions. We excluded the autonomy set from the factor estimations for these countries.

[13]The number of observations is smaller than in our preferred specifications because there are more missing values in the autonomy and policy set of questions.

[14]Relatedly, researchers could use our methodology to select existing questions to add to their own survey instruments (c.f. Crawfurd et al., 2024).

[15]We chose Prova Brasil for three reasons: first, it includes school identifiers that we can match to schools in the 2013 WMS sample (thereby allowing for an additional validation exercise); second, it can be combined with other rich datasets in Brazil to approach myriad research questions; and third, we use this index in an example application in Section 4.

between the management indices from Prova Brasil and the WMS scores for *the same school* (Table A.1 and Figure A.4, respectively). We include the details of this exercise in the Appendix for interested readers.

3.5 Limitations of our measurement approach

We have argued that our measurement approach using public data allows researchers to build and use management indices quickly and without large grants. We conclude this section with a discussion of limitations, and the steps that can be taken to address them.

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 national datasets (as is the case for our Prova Brasil-based index).

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.

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 need 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 or -99 is a common choice) and include a dummy variable that identifies the values which were imputed. This is done only with control variables; when data is missing for critical variables (say, outcome variables or key explanatory variables), we drop the observation.[16]

4 Application: unpacking turnover and test scores

So far, we have emphasized that management practices matter for organizational performance. Yet practices are (at the outset) implemented by *people* and it is well known that people matter too. To cite one particularly well-identified example, Huber et al. (2021) show that firms in Germany incurred significant losses when Jewish managers were forced out as a result of rising discrimination in Nazi Germany. A similar relationship can be found in schools: as we document below, there is a negative relationship between principal turnover and student performance in schools in Brazil.

It is natural to question *why* turnover matters. One possibility is that a new principal, deliberately or otherwise, de-stabilizes existing school management structures. Another is that turnover mechanically changes at least some principal characteristics — gender, age, experience, educational background — and these impact performance above and beyond management. Armed with our new measurement approach, it is straightforward to explore these issues. We do so in two stages. First, in Section 4.1 we use a regression discontinuity design (RDD) to test whether an exogenous 'turnover shock' changes school management practices. This is a key step in the causal chain illustrated in Figure 2, and evidence from a clean identification strategy is informative. Then, in Section 4.2, we expand beyond the RDD sample and use mediation analysis to quantify the importance of the two channels shown in Figure 2 — management versus managers — in mediating the relationship between turnover and test scores. Here, we follow the approach of Acharya et al. (2016), noting the necessary identifying assumptions as we go.

[16] We choose this simplest of methods because implementing alternative methods such as multiple imputation with datasets that have potentially many different survey weights becomes impractical.

4.1 A turnover shock and school management

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. We use the same turnover shock but instead examine the impact on school management practices.

We merge our Prova Brasil management index into the replication dataset (Moreira et al., 2021) and begin by reproducing the key Akhtari et al. (2022) results. To do so, we run their RD model and reproduce their RD graphs. This model is:

$$\begin{aligned} Y_{smt+1} =& \beta_0 + \beta_1 \ \mathbf{1}\{IVoteM_{mt} < 0\} + \beta_2 \ IVoteM_{mt} \\ &+ \beta_3 \ \mathbf{1}\{IVoteM_{mt} < 0\} \times IVoteM_{mt} + X^{'}_{smt} \ \beta_4 + \varepsilon_{smt}, \end{aligned} \tag{3}$$

where Y_{smt+1} is the school-level average test score for school s in municipality m constructed using the Prova Brasil survey wave from one year after the election at time t. The running variable is the incumbent vote margin, $IVoteM_{mt}$, defined as the vote share of the incumbent political party minus the vote share of the incumbent party's strongest opponent. The treatment variable is $\mathbf{1}\{IVoteM_{mt} < 0\}$, which takes the value one if the incumbent political party lost the election and zero if it won. The vector X_{smt} is a set of controls that includes the baseline test score prior to the election, school-level characteristics, and an election-cycle dummy. Standard errors are clustered at the municipality level.

For ease of exposition, we focus on the RD graphs, noting the RD estimate (and associated p-value) from estimation of equation (3) within the figure. The top two panels in Figure 3 compare student test scores in municipalities where the incumbent mayor *narrowly won* with student test scores in municipalities where the incumbent mayor *narrowly lost*, under the assumption that turnover in political party is as good as random in these close-election municipalities.[17] Panel (a) shows the main result on student test scores, and Panel (b) shows the placebo test for non-municipal schools.

Having replicated the Akhtari et al. (2022) results, we re-estimate equation (3) with

[17] Akhtari et al. (2022) find empirical support for this key identification assumption. We do not replicate all aspects of their work here.

Y_{smt+1} defined as the Prova Brasil management index from one year after the election at time t. All other variables remain the same, except that we control for the baseline school management score in place of the baseline school-level test score. The bottom two panels of Figure 3 report the RD graphs, and we include the regression tables in the Appendix (Table A.3). Panel (c) shows a clear discontinuity in management scores in municipal schools: the management index is *lower* in municipalities where the incumbent mayor narrowly loses an election (right side of the graph) than in municipalities where the incumbent mayor narrowly wins (left side of the graph). Panel (d) shows that this discontinuity is not present in non-municipal schools — an important placebo test since these schools would not have been subject to upheaval associated with political turnover at the municipal level.

To probe the premise in Akhtari et al. (2022) that political turnover triggered changes in school personnel, we split the sample by whether schools had, or had not, recently experienced a change in principal. Figure 4 shows evidence of a discontinuity in management scores in schools when there had recently been principal turnover but no evidence of a discontinuity when there had not (compare Panel a and b). Since this sub-sample approach departs from the clean quasi-experiment, we turn to a particular institutional arrangement for an additional test. 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 source of variation in principal turnover. The discontinuity in school management scores is present in the group of municipalities with an appointed structure (Panel c), but not in those where principals are civil servants (Panel d).

In summary, we have found evidence that an exogenous turnover shock changes school management practices. By splitting the RD sample, we have also found suggestive evidence that this operates via a change in school leadership, as hypothesized in the first horizontal arrow in Figure 2. Below, we explore the relative importance of this management practice channel, versus principal characteristics, in explaining why turnover matters for test scores.

4.2 Managers or management?

We begin by examining the correlation between school principal turnover and student test scores in our Prova Brasil data. To do so, we estimate the following regression model:

$$Y_{isr} = \beta_0 + \beta_1 \; T_{sr} + X_{sr}^{'} \; \beta_2 + \gamma_r + \varepsilon_{isr}, \tag{4}$$

where Y_{isr} is a test score outcome (averaged over math and portuguese) for student i in school s and region r in the Prova Brasil 2013 assessment.[18] The turnover variable, T_{sr}, is an indicator that takes the value one if the principal in school s was appointed in the previous two years and zero otherwise.[19] The vector X_{sr} includes school-level controls that are pre-treatment (determined prior to 2011), and γ_r is a region fixed effect. Standard errors are clustered at the school level. Our estimate of β_1 is reported in the first column of Table 5, and is negative: the partial effect of principal turnover is -0.06 SD of student learning.

Our aim is to quantify the importance of management practices versus principal characteristics, as measured in 2013, in mediating this relationship between turnover and test scores. Since both variables are post-treatment, adding them to equation (4) can introduce bias and is ill-advised (see e.g. Rosenbaum, 1984). Instead, we follow Acharya et al. (2016) and seek to estimate the *Average Controlled Direct Effect* (ACDE). This is the effect of changing the turnover variable while fixing the value of the mediator(s) at the same level for all units, averaged over all these units. By comparing our estimate of the ACDE with our estimate of β_1 from equation (4) we can gauge the importance of the mediator as a causal mechanism for the impact of turnover on test scores.[20]

The first step in this approach is to estimate the effect of the proposed mediator(s) on the test score outcome using the following model:

$$Y_{isr} = \alpha_0 + \alpha_1\ T_{sr} + M_{sr}^{'}\ \alpha_2 + X_{sr}^{'}\ \alpha_3 + Z_{isr}^{'}\ \alpha_4 + \mu_r + \epsilon_{isr}, \tag{5}$$

where Y_{isr} and T_{sr} are defined above, and u_r is a region fixed effect. The vector M_{sr} contains the proposed mediators. Depending on the specification, M_{sr} is either the 2013 Prova Brasil management score, or a set of principal characteristics as of 2013, or both. To justify the necessary identifying assumption (discussed below), we condition on pre-treatment school-level controls X_{sr} and a vector of intermediate confounders Z_{isr}. The latter vector includes a rich set of variables, measured in 2013, that could have been impacted by principal turnover and that may also be correlated with both the mediator and the test score outcome. The second step is to demediate the test score outcome: $\widetilde{Y}_{isr} = Y_{isr} - \hat{\alpha}_2 M_{sr}$, where $\hat{\alpha}_2$ is from equation (5). We then regress $\widetilde{Y}_{isr}$ on T_{sr}, X_{sr}, Z_{isr} and μ_r. The coefficient on the turnover indicator

[18] Throughout this section, we focus on a single wave for transparency in determining pre- and post-treatment variables. Results for Prova Brasil 2011 and 2015 are provided in the Appendix.

[19] We follow the mediation literature in referring to the variable whose effects we wish to unpack as the 'treatment.' This is not meant to imply a belief that turnover was randomly assigned.

[20] Formally, the difference in these coefficients is the sum of two terms: (i) the average natural indirect effect (the effect of turnover on test scores operating only through the mediator) and (ii) an interaction effect that captures how the mediator influences the direct effect of treatment.

in this regression is our estimate of the ACDE. Standard errors are from nonparametric bootstrap (with 1000 repetitions) to account for the first stage estimation.

This two-step procedure for estimation of the ACDE needs two identifying assumptions: (1) sequential unconfoundedness, and (2) no intermediate interactions. The former requires that the relevant assumptions (no omitted variables) hold to enable estimation of the effect of turnover on test scores and, separately, the effect of the mediator(s) on test scores, conditional on T_{sr}, X_{sr} , and Z_{isr}. We include a wealth of variables in X_{sr} and Z_{isr} which makes this plausible but it is, ultimately, still an assumption. No intermediate interactions requires that the effect of the mediator on the test score outcome is independent of the intermediate confounders in Z_{isr}. As Acharya et al. (2016) note, this is similar to omitting an interaction term from a regression model; if it fails, estimates will be weighted averages of ACDEs within levels of the intermediate confounders.

The results of this exercise are reported in Table 5. Each of columns (2) to (4) fixes a different set of mediators.[21] Figure 5 shows the coefficients and 95 percent confidence intervals graphically. The first row reports the original 'treatment' effect of school principal turnover on student test scores (average of math and portuguese) in the Prova Brasil 2013 assessment. The next row reports the ACDE of principal turnover, fixing school management score as a mediator. We see that management accounts for 22 percent of the original effect.[22] The following row reports the ACDE of principal turnover, fixing principal characteristics as mediators. This set of mediators accounts for 4 percent of the original effect. The final row combines all mediators and shows that together they account for 26 percent of the original effect. The results in the Appendix, for the 2011 and 2015 Prova Brasil survey waves, echo this finding. In all three years, management and (to a lesser extent) managers are important mediators of the negative relationship between turnover and test scores, although a sizeable direct effect (the dashed arrow in Figure 2) always remains.

Our measurement approach has enabled us to make progress exploring why turnover matters for test scores. The conclusion that we draw from our different empirical analyses is that turnover matters, at least in part, through a de-stabilizing effect that reduces the quality of school management practices which, in turn, lowers student test scores. The framework of

[21] Potential mediating variables that are not included in M_{sr} in a given specification are instead included in Z_{isr}. This is particularly important for principal characteristics because, as Figure 2 hypothesizes, these variables could impact both management practices and student test sores.

[22] We have performed a sensitivity analysis (with management score as the sole mediator) to explore the implications of the sequential unconfoundedness assumption, applying the method in Acharya et al. (2016). In the most plausible case of unobserved confounders that are positively correlated with both management and test scores, our estimate of the ACDE is biased downwards, meaning that we would *understate* the mediating role of management.

Section 2, which is supported by the data, suggests mechanisms for this management effect that operate via selection and incentives of the actors within the school.[23]

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, an important 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.

This paper responds to these challenges and makes three distinct contributions. The first is the development of a new approach to measurement at scale using existing public data sources. Both of our new school management indices (PISA and Prova Brasil) 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 that proposes potential channels through which management structures could impact student learning: selection and incentives of the actors within the school. We find empirical evidence to support these mechanisms in both the PISA and Prova Brasil data. The third contribution, where we apply our new measurement approach and the theoretical framework, is to unpack the negative relationship between managerial turnover and organizational performance (Akhtari et al., 2022; Huber et al., 2021, and Table 5).

An important implication of our results is that both people and practices matter for school performance. 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 implement 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 instill but recent research suggests it can be effective to teach managers how to adopt better practice (e.g. Anand et al., 2023; Beg et al., 2023; Bloom et al., 2013; Roland G. Fryer, 2017).

[23]In Section 3, we noted that Table 2 (using PISA data) supports the four predictions from the theory. Table A.2 in the Appendix presents very similar results using data from Prova Brasil.

References

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review 110*(1), 512–528.

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., 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.

Brown, C. and T. Andrabi (2025). Inducing positive sorting through performance pay: Experimental evidence from Pakistani schools. Working paper, Chicago.

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.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

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.

Fuchs, T. and L. Wössmann (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics 32*, 433–464.

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 (2021). Discrimination, managers, and firm performance: Evidence from "Aryanizations" in Nazi Germany. *Journal of Political Economy 129*(9), 2455–2503.

Jackson, C. K. (2025, September). When decentralization works: Leadership, local needs, and student achievement. Working Paper 34216, National Bureau of Economic Research.

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 (2024). 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*.

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.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society A 147*(5), 656–666.

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., S. Ohlmacher, J. Van Reenen, M. Bennedsen, N. Bloom, A. Choudhary, L. Foster, J. Groenewegen, A. Grover, S. Hardeman, L. Iacovone, R. Kambayashi, M. C. Laible, R. Lemos, H. Li, A. Linarello, M. Maliranta, D. Medvedev, C. Meng, J. Miles

Touya, N. Mandirola, R. Ohlsbom, A. Ohyama, M. Patnaik, M. Pereira-López, R. Sadun, T. Senga, F. Qian, and F. Zimmermann (2024). The international empirics of management. *Proceedings of the National Academy of Sciences 121*(45), e2412205121.

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.

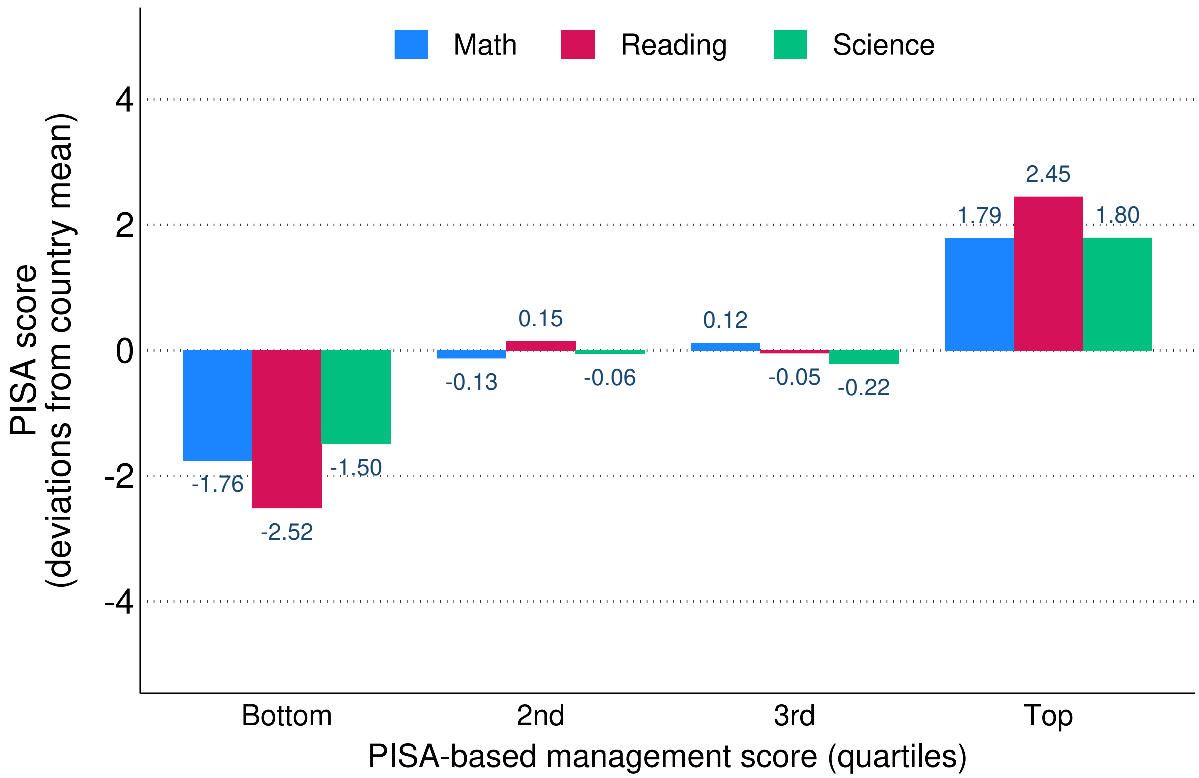
Wössmann, L. (2005). The effect heterogeneity of central examinations: Evidence from TIMSS, TIMSSRepeat and PISA. *Education Economics 13*(2), 143–169.

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: Index validation: management and student performance



Note: Data is from PISA 2012 including 15,192 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.

Figure 2: Principal turnover and test scores, stylized causal chain

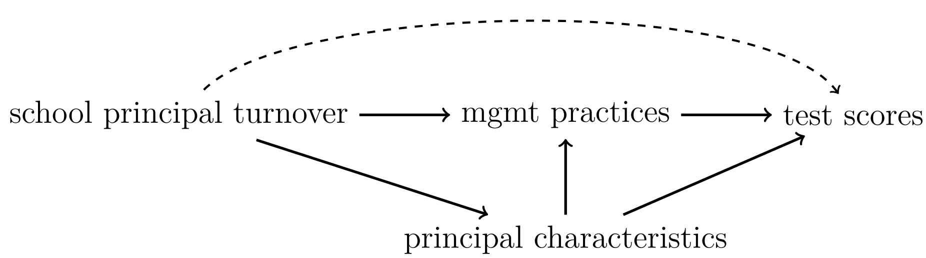
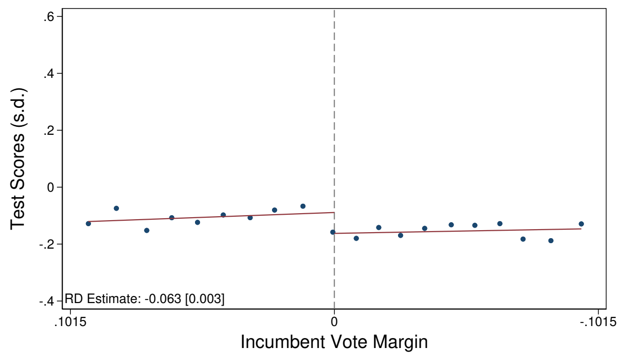
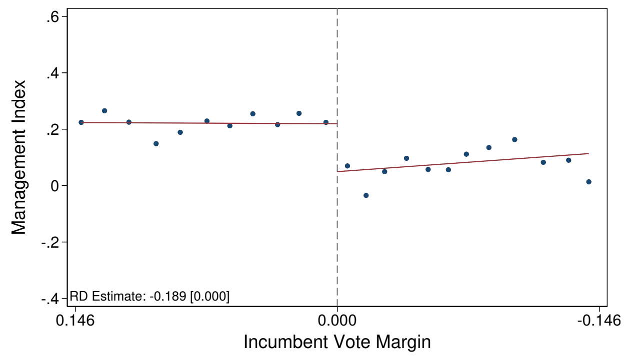


Figure 3: Political turnover, test scores, and school management

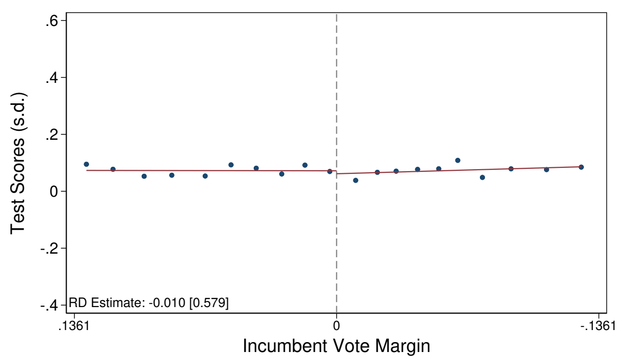
(a) Test Scores: municipal (treated) schools



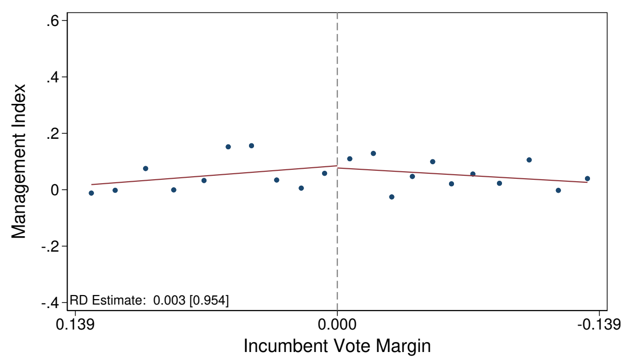
(b) Test Scores: non-municipal (placebo) schools



(c) Management: municipal (treated) schools



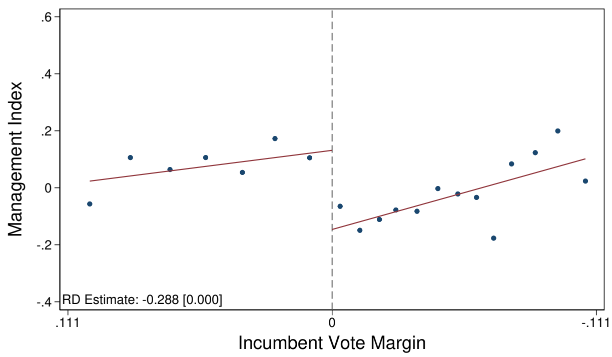
(d) Mgmt: non-municipal (placebo) schools



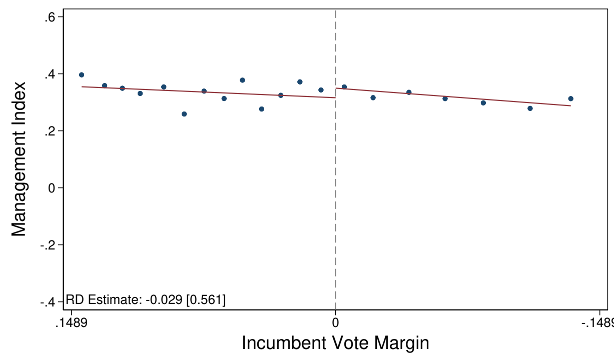
Note: Panel (a) replicates Akhtari et al. (2022) using the replication files from the American Economic Review. It shows the average of individual-level test scores by bins of $IVoteM_{mt}$ in municipal schools, pooling students from grade 5 and grade 9 and controlling for the average, school-level test scores for the respective grade at baseline. Municipalities with $IVoteM_{mt} < 0$ experienced a change in the political party of the mayor. Municipalities with $IVoteM_{mt} > 0$ 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 (c) repeats the analysis but with the average standardized management score in municipal schools by bins of $IVoteM_{mt}$ as the outcome variable, controlling for the standardized management score in the baseline year (year before the election). Panels (b) and (d) repeat the analysis for for non-municipal schools (a placebo test).

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

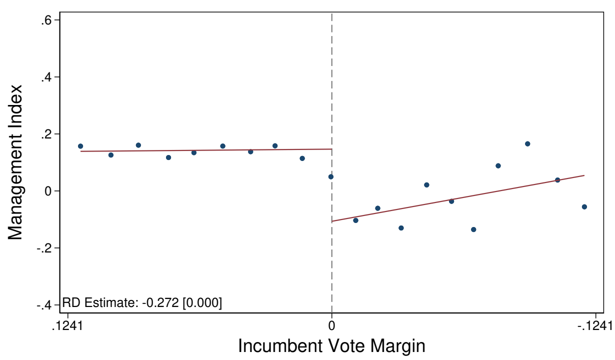
(a) New principal



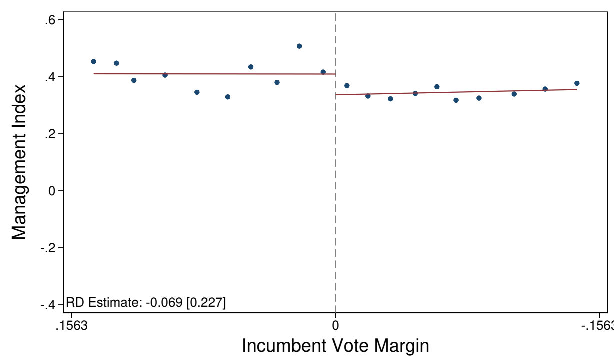
(b) No principal turnover



(c) Principals are politically appointed

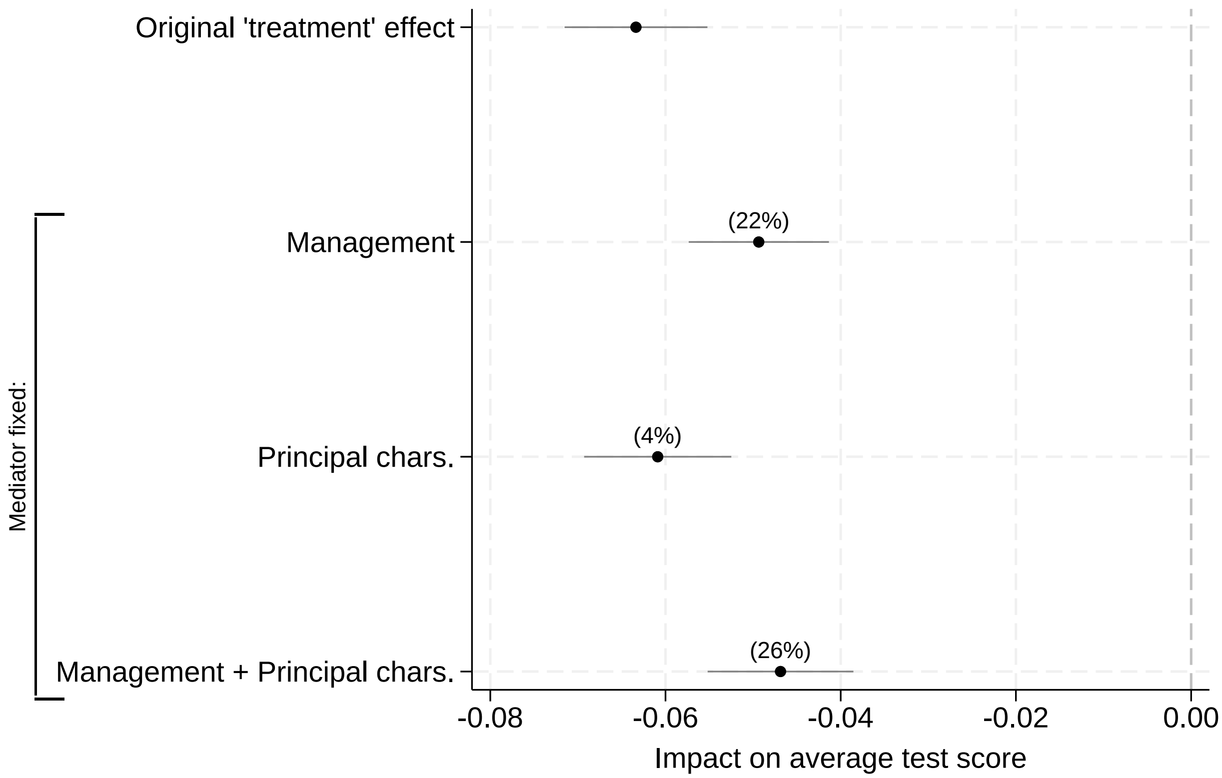


(d) Principals are civil servants



Note: See note to Figure 3 for details on the methodology. All panels use the average standardized management score as the outcome variable, controlling for the standardized management score in the baseline year (year before the election). The top two panels split the sample using a question from the Prova Brasil school principal survey asking whether the principal is newly appointed (within the last two years). Panel (a) includes schools where there is a new principal, and Panel (b) includes schools where the principal has not changed in the last two years. The bottom two panels split the sample by the system of political appointment for school principals in each municipality. Panel (c) includes schools where principals are political appointees, and Panel (d) includes schools where school principals are civil servants.

Figure 5: Mediation analysis: principal turnover



Note: Data from Prova Brasil. The top plot shows the original 'treatment' effect of school principal turnover on student test scores (in the Prova Brasil 2013 assessment). This coefficient and its 95 percent confidence interval are obtained from estimating equation (4). The second plot is the Average Controlled Direct Effect (ACDE) of turnover on test scores. The figure in brackets is the percent drop in the estimated treatment effect when school management score is fixed; it captures the share of the overall treatment effect on test scores attributable to this mediator. The third plot is the ACDE using a set of principal characteristics (binary indicators for gender, age over 50 years, experience over 15 years, wage over 7 times the minimum, and only one job, plus a set of education dummies) as mediators. The fourth plot combines all mediators. We estimate the ACDE using the techniques described in Acharya et al. (2016).

Table 1: Index validation: management and student performance

	Reading PISA Points			Math PISA Points			Science PISA Points		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Management Index	3.526***	2.399**	2.224***	3.015***	2.110**	1.857**	2.761***	1.950*	1.749**
	(1.018)	(0.987)	(0.814)	(0.993)	(1.010)	(0.806)	(1.001)	(1.023)	(0.815)
R-Squared	0.243	0.290	0.423	0.307	0.342	0.450	0.299	0.330	0.430
# Students	410,732	410,732	410,732	410,732	410,732	410,732	410,732	410,732	410,732
# Schools	15,192	15,192	15,192	15,192	15,192	15,192	15,192	15,192	15,192
Mfx (in months of learning)	1.06	0.72	0.67	0.90	0.63	0.56	0.83	0.58	0.52
Controls									
Country FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
School		Y	Y		Y	Y		Y	Y
Student			Y			Y			Y

Note: Data from PISA 2012. Standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively. 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 constructed management index standardized using the overall distribution. All specifications include 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). 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: Index validation: management and school functioning

Outcome variable:	Teacher shortage		Teacher motivation		Teacher effort		Household effort	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Management Index	-0.058**	-0.058***	0.303***	0.320***	0.047*	0.069***	0.260***	0.287***
	(0.023)	(0.022)	(0.027)	(0.026)	(0.025)	(0.024)	(0.031)	(0.029)
R-Squared	0.030	0.049	0.088	0.119	0.008	0.046	0.078	0.155
# Schools	12,144	12,144	12,144	12,144	12,144	12,144	12,144	12,144
Controls								
Country FE	Y	Y	Y	Y	Y	Y	Y	Y
School		Y		Y		Y		Y
Student		Y		Y		Y		Y

Note: Data from PISA 2012. Standard errors in parentheses. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively. The table reports coefficients from school-level regressions of the constructed management index standardized using the overall distribution on each of the standardized indices of school functioning. 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). 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: Index robustness: decision nodes and index method

	Decision Nodes				Index Method		
	(1)	(2)	(4)	(3)	(5)	(6)	(7)
Panel A: Reading							
Management Index	2.224***	1.869**	2.143***	1.807**	1.968***	1.952**	2.299**
	(0.814)	(0.808)	(0.823)	(0.774)	(0.762)	(0.793)	(0.919)
R-Squared	0.423	0.423	0.420	0.408	0.423	0.422	0.423
Panel B: Math							
Management Index	1.857**	1.478*	2.016**	1.610**	1.402*	1.468*	2.158**
	(0.806)	(0.794)	(0.821)	(0.819)	(0.775)	(0.768)	(0.944)
R-Squared	0.450	0.449	0.444	0.439	0.450	0.449	0.450
Panel C: Science							
Management Index	1.749**	1.377*	1.641**	1.454*	1.513**	1.383*	1.946**
	(0.815)	(0.800)	(0.836)	(0.803)	(0.763)	(0.767)	(0.923)
R-Squared	0.430	0.430	0.428	0.416	0.430	0.429	0.431
# Students	410,732	410,200	391,999	388,701	410,732	408,528	410,200
# Schools	15,192	15,176	14,526	14,342	15,192	15,119	15,176
Controls							
Country FE	Y	Y	Y	Y	Y	Y	Y
School	Y	Y	Y	Y	Y	Y	Y
Student	Y	Y	Y	Y	Y	Y	Y
Index	Core mgmt	+ mgmt broad	+ autonomy	+ policy	Core mgmt	Core mgmt	Core mgmt
Nodes included	all	1,2,3	1,2,4	1,3,4	all	all	all
Method	Avg Practice	Avg Practice	Avg Practice	Avg Practice	Avg Questions	PCA	Anderson (2008)

Note: Data from PISA 2012. Each cell is a regression using OECDs `repest` Stata command and the plausible values of test scores for reading (Panel A), math (Panel B) and science (Panel C). The first four columns iterate over different sets of variables included in the management index. The last three columns iterate over methods of building indices using the same set of variables as Column (1). All columns include country fixed effects, school and student controls as in Table 1.

Table 4: Index robustness: exploratory factor analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Reading						
Factor Index	0.727***	-1.048***	0.256	0.611	-0.465	-0.872
	(0.272)	(0.369)	(0.373)	(0.486)	(0.507)	(0.557)
R-Squared	0.416	0.416	0.415	0.415	0.415	0.416
Panel B: Math						
Factor Index	0.717***	-1.017***	0.108	0.781	-0.472	-0.778
	(0.245)	(0.359)	(0.432)	(0.493)	(0.519)	(0.527)
R-Squared	0.448	0.448	0.447	0.447	0.447	0.447
Panel C: Science						
Factor Index	0.642**	-0.986***	0.079	0.631	-0.215	-0.670
	(0.258)	(0.340)	(0.391)	(0.509)	(0.505)	(0.534)
R-Squared	0.425	0.425	0.424	0.425	0.424	0.425
# Students	290,178	290,178	290,178	290,178	290,178	290,178
# Schools	10,663	10,663	10,663	10,663	10,663	10,663
Latent factor	1	2	3	4	5	6
Controls						
Country FE	Y	Y	Y	Y	Y	Y
School	Y	Y	Y	Y	Y	Y
Student	Y	Y	Y	Y	Y	Y

Note: Data from PISA 2012. Each cell is a regression using OECDs `repest` Stata command and the plausible values of test scores for reading (Panel A), math (Panel B) and science (Panel C). All columns include country fixed effects, school and student controls as in Table 1. The main explanatory variable, Factor Index, refers to the latent factors identified by Exploratory Factor Analysis. Each column iterates over the six key factors identified through the analysis as the main explanatory variable.

Table 5: Mediation analysis: Principal turnover and student test scores

	Original outcome	De-mediated outcome		
	(1)	(2)	(3)	(4)
Treatment variable				
New Principal = 1	-0.063***	-0.049***	-0.061***	-0.047***
	(0.004)	(0.004)	(0.004)	(0.006)
Pre-treatment variables				
Urban = 1	Y	Y	Y	Y
Education spending (2011)	Y	Y	Y	Y
Financial controls (2011)	Y	Y	Y	Y
School characteristics (2011)	Y	Y	Y	Y
# Students	3,668,446	3,668,446	3,668,446	3,668,446
# Schools	47,878	47,878	47,878	47,878
Variables used in first state of mediation analysis				
Mediators				
Management score		Y		Y
Principal characteristics			Y	Y
Post-treatment controls				
Education spending (2013)		Y	Y	Y
Financial controls (2013)		Y	Y	Y
School characteristics (2013)		Y	Y	Y

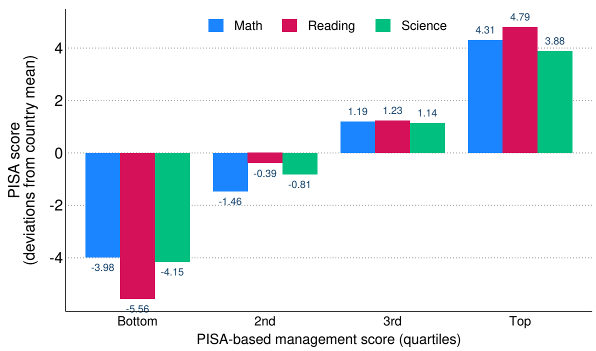
Note: Standard errors in parentheses. *** denotes significance at the 1 percent level. Column (1) reports results from the student-level specification in equation (4). The outcome variable is student test score (averaged over math and portuguese) in the 2013 Prova Brasil assessment. The treatment variable is an indicator that takes the value 1 if the school prinicpal is new, having been appointed between 2011 and 2013, and 0 otherwise. **Pre-treatment variables**: indicator for school location; total spent on education by the municipal government in the school's municipality in 2011; a set of indicators for the school's receipt of financial transfers in 2011 from each of the federal, state and municipal government; and school characteristics in 2011 (student-teacher ratio, log of the number of students, indicators for whether the school has a computer lab and internet access, and an infrastructure index). Columns (2) to (4) report results from the mediation analysis. The first stage (not reported) is estimation of equation (5). Column (2) uses the 2013 Prova Brasil management index as a mediator. Column (3) uses a set of 2013 principal characteristics (binary indicators for gender, age over 50 years, experience over 15 years, wage over 7 times the minimum, and only one job, plus a set of education dummies) as mediators. In both columns, the variable that is not used as a mediator is instead included as a post-treatment control. Column (4) uses all mediators. In all three columns, **post-treatment variables** are: education spending, financial controls and school characteristics as of 2013, plus teacher characteristics in 2013 aggregated to school-level (gender, ethnicity, education, tenure) and student characteristics in 2013 (age, gender, socio-economic status, ethnicity, work status, parental literacy). The columns report results from regressing the *demediated outcome* on the treatment variable and the same set of controls used in column (1). The coefficient on the treatment variable can be interpreted as the Average Controlled Direct Effect of turnover on test scores (Acharya et al., 2016).

ONLINE APPENDIX—NOT FOR PUBLICATION, December 2025
"Understanding school management with public data: A new measurement approach and applications, Leaver, Lemos and Scur"

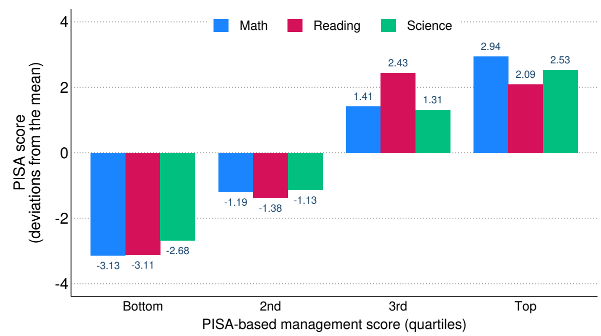
A Additional Figures and Tables

Figure A.1: Index robustness: management and student performance, alternative samples

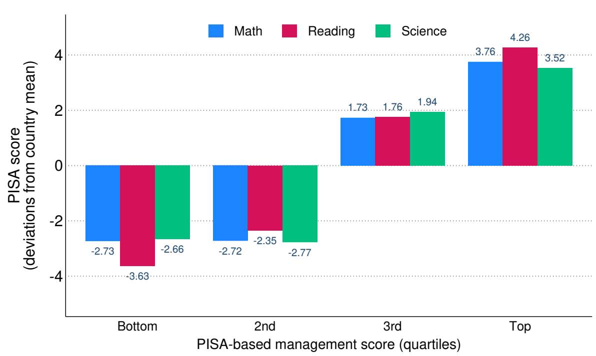
(a) PISA 2012, common questions with 2022



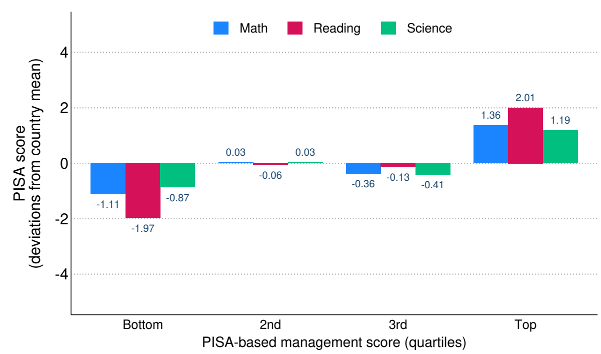
(b) PISA 2022



(c) PISA 2012, Anderson (2008) Index

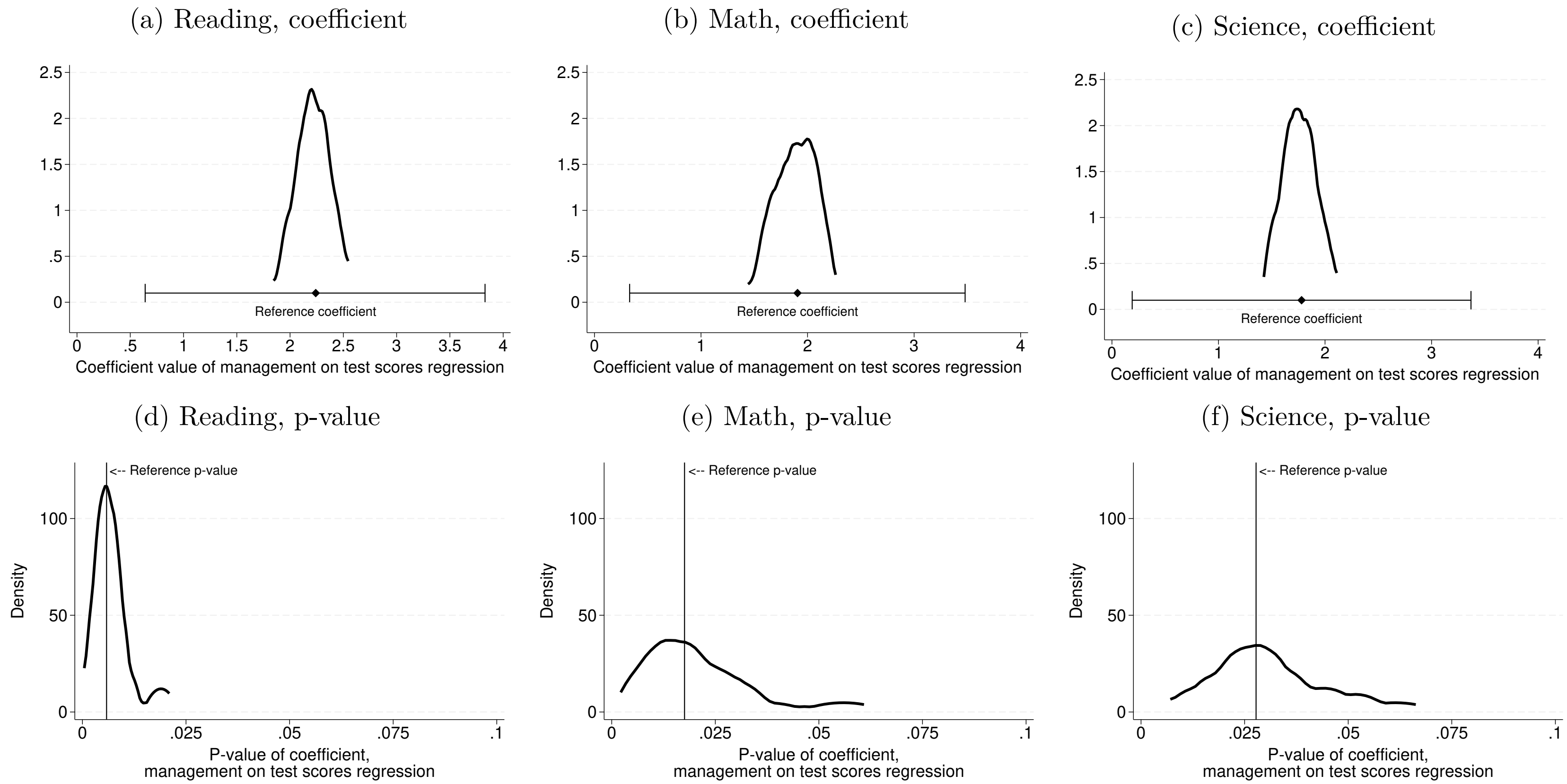


(d) PISA 2012, Principal Component Analysis



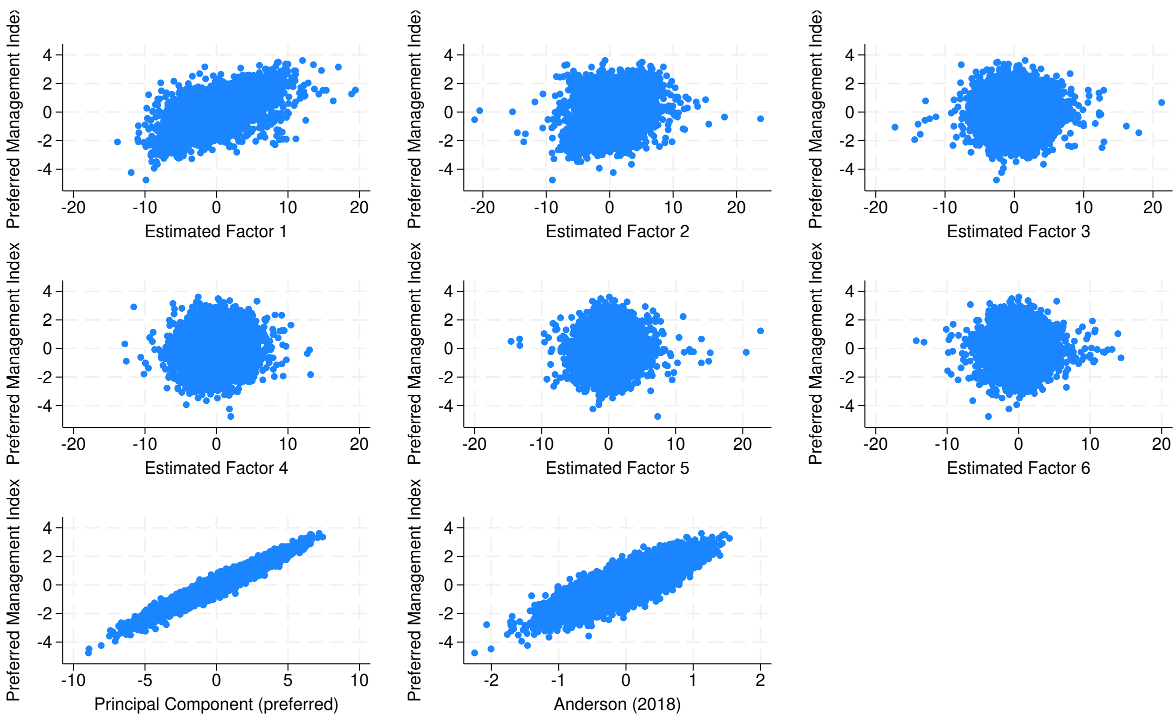
Note: Data from PISA 2012 and 2022. Panels (a), (c) and (d) use data from PISA 2012. Panel (b) uses data from PISA 2022. Student outcomes are estimated using five plausible values and collapsed at the school level using PISA's senate weights. Quartiles of the management index are built at the country level, for respective years. Test scores are presented as deviations from the subject-specific country mean for respective years. Panel (a) uses the set of common questions from the principal questionnaires in PISA 2012 and 2022 that pass the four decision nodes. Panel (b) uses the common set of questions in the 2022 questionnaire. Panel (c) builds the main management index using the Anderson (2008) methodology, where greater weight is given to variables with greater variation. Panel (d) builds the main management index using standard Principal Component Analysis methodology.

Figure A.2: Index robustness: leave-one-out exercise



Note: Data from PISA 2012. The top graphs (a,b,c) plot the distribution of the coefficients of $MGMT_{sc}$ when running Equation 2 for indices that exclude one of the management practices one at a time from the index construction ('leave one out' exercise). The marker and line at the bottom show the preferred specification 'benchmark' coefficient and confidence interval. The bottom panels plot the distribution of the p-values of the estimated coefficients. The vertical line denotes the 'benchmark' specification p-value of the management coefficient as a reference point.

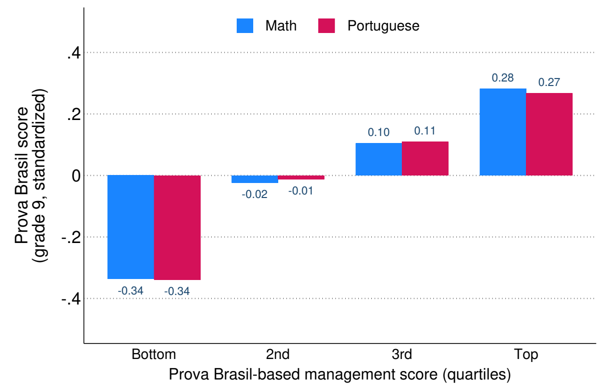
Figure A.3: Index robustness: exploratory factor analysis comparison



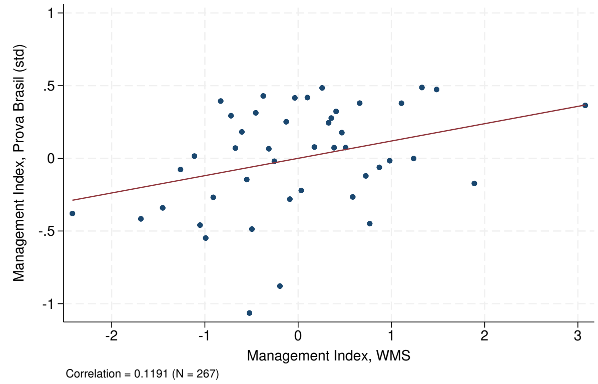
Note: Data from PISA 2012. Each of the first six plots on the top two rows compare the 'preferred management index' to an index built out of the first six factors extracted from an exploratory factor analysis exercise. The last two plots on the bottom row compare the preferred index to indices built using the same set of questions but different index methodologies: PCA and Anderson (2008).

Figure A.4: Index validation: management and student performance, alternative dataset (Prova Brasil)

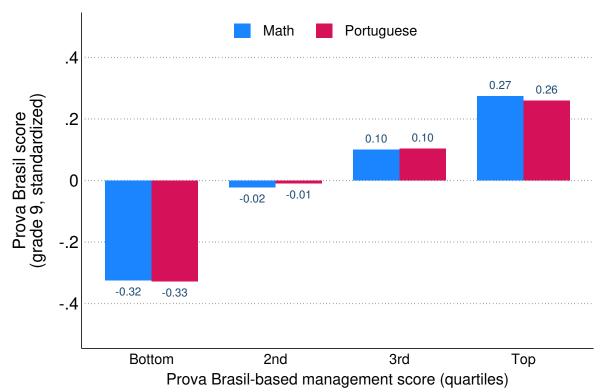
(a) Preferred Index: test scores by quartile



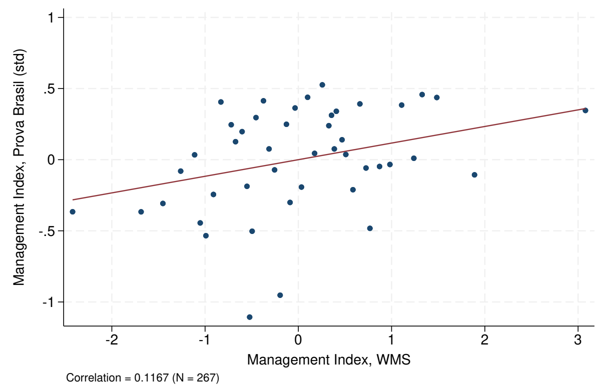
(b) Preferred Index: school-specific scores



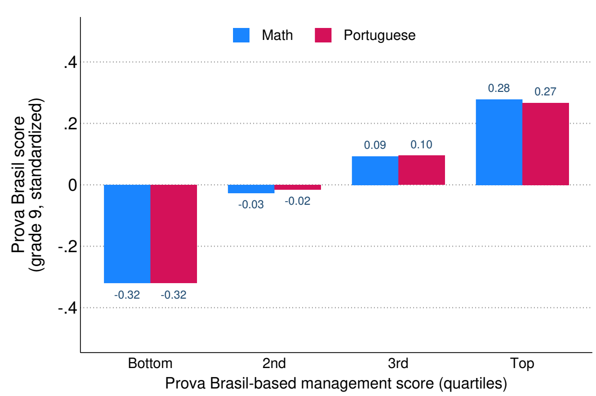
(c) Anderson (2008) Index: test scores by quartile



(d) Anderson (2008) Index: school-specific scores



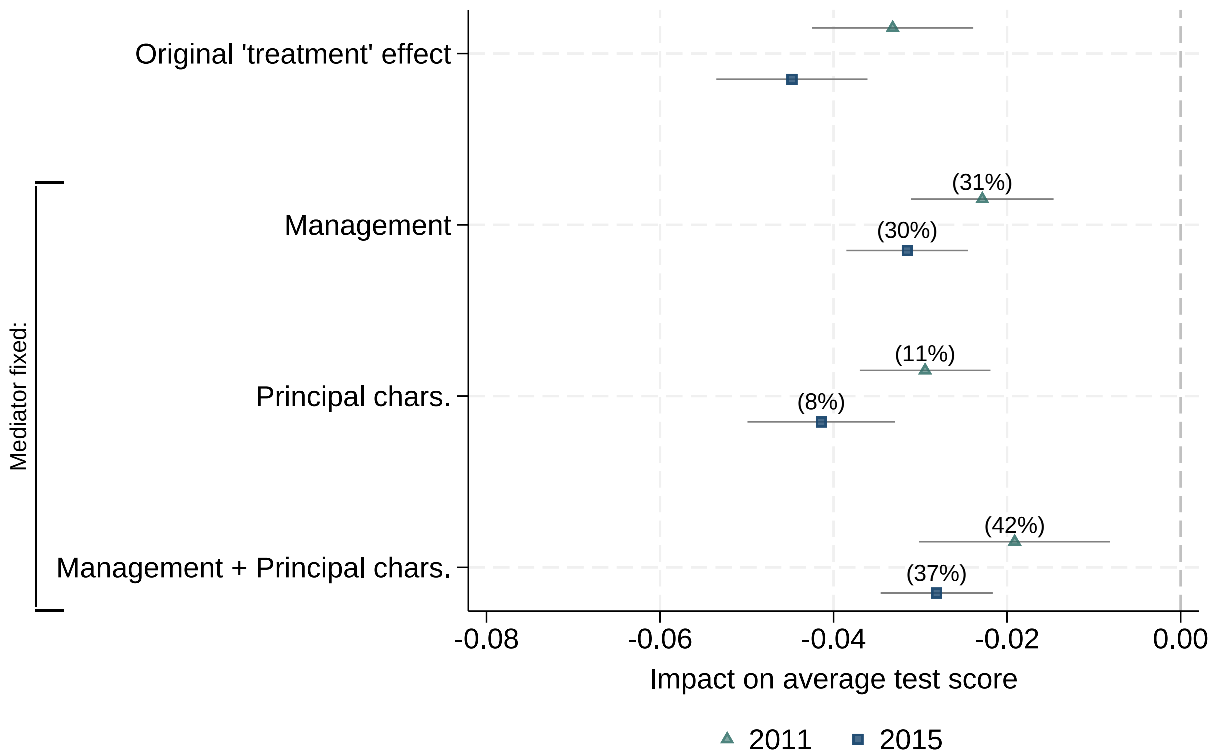
(e) Principal Component Analysis: test scores by quartile



(f) Principal Component Analysis Index: school-specific scores

Note: Data is from Prova Brasil (2013), including 32,573 schools. Sample restricted to schools with grade 9 to maintain closer comparability to the WMS sample. Panels (a), (c) and (e) show test scores in deviations from the subject-specific mean, by management quartile. Panel (a) uses our preferred index building methodology, following the WMS. Panel (c) builds the main management index using the Anderson (2008) methodology, where greater weight is given to variables with greater variation. Panel (e) builds the main management index using standard Principal Component Analysis methodology. Panels (b), (d) and (f) also use data from the World Management Survey. In these panels, the sample contains schools which have data for both Prova Brasil and WMS in 2013 (267 schools), matched at the school level via school identifiers. All graphs are binned scatter plots using 45 quantiles.

Figure A.5: Mediation analysis: 2011 and 2015



Note: This figure replicates Figure 5 using the Prova Brasil 2011 and 2015 survey waves. The top plot shows the original 'treatment' effect of school principal turnover on student test scores (in either the Prova Brasil 2011 or 2015 assessment). This coefficient and its 95 percent confidence interval are obtained from estimating equation (4), with variables defined appropriately for 2011 and 2015. The second plot is the Average Controlled Direct Effect (ACDE) of turnover on test scores using management as a mediator. The third plot is the ACDE using a set of principal characteristics (binary indicators for gender, age over 50 years, experience over 15 years, wage over 7 times the minimum, and only one job, plus a set of education dummies) as mediators. The fourth plot combines all mediators. We estimate the ACDE using the techniques described in Acharya et al. (2016).

Table A.1: Index validation: management and student performance, alternative dataset (Prova Brasil)

	Portuguese Score				Mathematics Score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Management Index	0.147***	0.069***	0.044***	0.018***	0.162***	0.073***	0.050***	0.019***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
R-Squared	0.018	0.068	0.160	0.234	0.022	0.085	0.151	0.245
# Students	12,263,694	12,263,694	12,263,694	12,263,694	12,263,694	12,263,694	12,263,694	12,263,694
# Schools	68,704	68,704	68,704	68,704	68,704	68,704	68,704	68,704
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
State FE		Y	Y			Y	Y	
Controls	Y	Y	Y	Y	Y	Y	Y	Y
School FE				Y				Y

Note: Data from Prova Brasil. OLS regressions were run with the student-level dataset, pooling grades 5 and 9, for years 2011 to 2015. Standard errors clustered at the school level. Test scores are normalized within grade. All specifications include year fixed effects. **Student controls** include: gender, student households consumption index, a set of dummies for race, dummies for mother educational attainment (grades 1-5, grades 6-9, secondary grades 10-12, and college). **School controls** include: 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, 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, class-year-level share of white teachers, share of teachers holding a college degree, and average teacher tenure.

Table A.2: Index validation: management and school functioning, alternative dataset (Prova Brasil)

Outcome variable:	Teacher shortage		Teacher motivation		Teacher effort		Household effort	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Management Index	-0.086***	-0.112***	0.347***	0.365***	0.074***	0.041***	0.123***	0.047***
	(0.003)	(0.005)	(0.003)	(0.006)	(0.003)	(0.005)	(0.003)	(0.004)
R-Squared	0.007	0.609	0.119	0.575	0.006	0.657	0.016	0.703
# School-year obs	169,557	169,557	167,586	167,586	169,601	169,601	169,622	169,622
# Schools	68,889	68,889	68,419	68,419	68,898	68,898	68,912	68,912
Controls								
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y				Y		Y
School		Y		Y		Y		Y
Student		Y		Y		Y		Y

Note: Data from Prova Brasil. OLS regressions were run with the student-level dataset, pooling grades 5 and 9, for years 2011 to 2015. Standard errors clustered at the school level. Test scores are normalized within grade. All specifications include year fixed effects. **Student controls** include: gender, student households consumption index, a set of dummies for race, dummies for mother educational attainment (grades 1-5, grades 6-9, secondary grades 10-12, and college). **School controls** include: 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, 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, class-year-level share of white teachers, share of teachers holding a college degree, and average teacher tenure.

Table A.3: Political turnover and school management scores

	Outcome: Management Index					
	(1)	(2)	(3)	(4)	(5)	(6)
Municipal Schools						
$1[IVoteM\_mt < 0]$	-0.189***	-0.185***	-0.213***	-0.207***	-0.247***	-0.238***
	(0.038)	(0.037)	(0.054)	(0.052)	(0.043)	(0.042)
Baseline Management Index	0.362***	0.321***	0.353***	0.310***	0.352***	0.311***
	(0.011)	(0.010)	(0.013)	(0.013)	(0.012)	(0.011)
R-Squared	0.166	0.184	0.163	0.184	0.160	0.177
# School-year obs	11,607	11,607	6,117	6,117	9,080	9,080
# Schools	2,516	2,516	1,563	1,563	2,130	2,130
Using Bandwidth	0.146	0.146	0.070	0.070	0.110	0.110
Optimal Bandwidth	0.146	0.146	0.146	0.146	0.146	0.146
Non-Municipal Schools						
$1[IVoteM_{mt} < 0]$	0.003	0.004	0.058	0.082	0.009	0.016
	(0.054)	(0.051)	(0.068)	(0.064)	(0.055)	(0.053)
Baseline Management Index	0.418***	0.384***	0.400***	0.363***	0.402***	0.369***
	(0.017)	(0.015)	(0.019)	(0.017)	(0.015)	(0.014)
R-Squared	0.191	0.211	0.179	0.203	0.180	0.202
# School-year obs	6,992	6,992	3,965	3,965	5,663	5,663
# Schools	2,183	2,183	1,390	1,390	1,875	1,875
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
Controls						
School controls		Y		Y		Y

Note: Standard errors in parentheses. The top panel reports the coefficient on political party turnover from the regression in equation (4) using the sample of municipal schools. In all specifications, we control for baseline standardized management scores in the year before the election. Additional controls are a set of school-level indicators (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). The bottom panel repeats the analysis using non-municipal schools (state and federal schools). Only public schools participate in the Prova Brasil exam.

B Appendix: Theoretical Framework

B.1 Model

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 τ 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 Δ 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 γ 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 θ denotes the teacher's *ability.* This can also be thought of as the realization of a random variable, assumed to be independent of τ. The teacher observes this realization perfectly, while (at the time of hiring) employers observe nothing. Draws of the error term ε are independent across employments. We assume throughout that ε is mean zero and distributed $U[\underline{\varepsilon}, \bar{\varepsilon}]$.

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 β 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 Δ. 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 γ. 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, Δ, base pay, and γ 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 (τ, θ) 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.

B.2 Analysis

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 τ 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 τ and ability θ 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 ε, 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

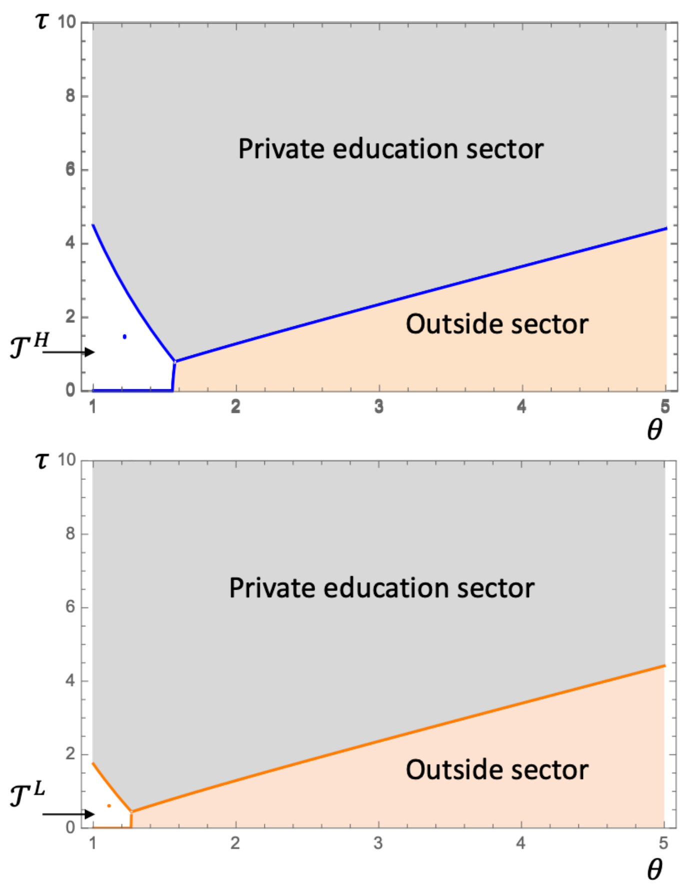
$$P^O = \Pr\left(\theta\ e + \varepsilon^O > \overline{z}\right) = \Pr\left(\theta\ e - \overline{z} > -\varepsilon^O\right) = \frac{\overline{\varepsilon} + \theta\ e - \overline{z}}{\overline{\varepsilon} - \underline{\varepsilon}}.$$

The first order condition for this optimization problem is

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

which yields the solution stated above. □

Figure B.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 $(1.11, 0.60)$.

We now use these effort levels to study teacher selection. Figure B.1 is based on a numerical example, as described in the figure note.[24] The unshaded region in the top panel of the figure shows $\mathcal{T}^H$, the set of (θ, τ) 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 τ_P^H traces out the loci of (θ, τ) 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

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

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 τ_P^H stated above. Fixing θ, 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 τ_O^H traces out the loci of (θ, τ) 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

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

Substituting for e^H and e^O from Lemma 1, together with the parameters in the numerical example, and rearranging for τ yields the expression for τ_O^H stated above. Fixing θ, 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 (θ, τ) 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 τ_P^L traces out the loci of (θ, τ) types who, anticipating subsequent teacher and household effort, are indifferent between accepting the job in the assigned low management public school

[24] Calculations were performed in Mathematica; the notebook file is available on request.

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 τ_P^L stated above. Fixing θ, 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 τ_O^L traces out the loci of (θ, τ) 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 τ yields the expression for τ_O^L stated above. Fixing θ, 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.

C Variable classification: PISA and Prova Brasil