



How Much Do Principals Matter When Politicians Choose Them? Evidence from Dated Successions and Close Elections in Brazil

Marco Cauduro

Goldenbird Capital

Estimates of how much school principals matter differ by an order of magnitude, in part because conventional estimators require principal identifiers and absorb sampling noise into the estimated dispersion. We identify the variance of principal-by-school effects without identifiers, using successions dated by principals' reported tenure in Brazil's 2011 national assessment, which imply known probabilities of a change between consecutive tests. The standard deviation is 0.05 student standard deviations in grade 5 and indistinguishable from zero in grade 9, against 0.18 for plug-in estimates on the same data. Dispersion is concentrated in networks where principals are appointed rather than competitively selected. These estimates imply that principal replacement can explain only a small share of the learning losses that prior work attributes to mayoral party turnover. In close elections, a change in the ruling party replaces municipal principals and, where these posts are discretionary, also disrupts the teaching staff and worsens grade progression. The rules by which principals are chosen matter more than the identity of any individual principal.

VERSION: October 2026

Suggested citation: Cauduro, Marco. (2026). How Much Do Principals Matter When Politicians Choose Them? Evidence from Dated Successions and Close Elections in Brazil. (EdWorkingPaper: 26-1618). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/953t-2q19>

How Much Do Principals Matter When Politicians Choose Them?

Evidence from Dated Successions and Close Elections in Brazil

Marco Cauduro*

September 2026

Abstract

Estimates of how much school principals matter vary by an order of magnitude, partly because standard estimators require principal identifiers and confound noise with dispersion. Dating successions from tenure reports in Brazil's national assessment, we identify the variance of principal-by-school effects from excess test-score volatility around successions. Its standard deviation is 0.05 student standard deviations in grade 5 and near zero in grade 9, and it is concentrated where principals are appointed. These estimates imply that principal replacement explains little of the learning losses attributed to mayoral party turnover. In close elections, turnover in appointment networks disrupts teaching staff.

Keywords: school principals; manager effects; limited mobility bias; patronage; political turnover; regression discontinuity; Brazil.

JEL codes: I21, I28, H75, D73, M54.

*Goldenbird Capital. Email: mac@goldenbird.capital. This paper builds on the author's doctoral dissertation at FGV EAESP (Cauduro 2018). I thank Ingrid, Gisele and Felipe. All data are public; replication code is described in the Data Availability statement. All errors are my own.

1 Introduction

School principals occupy a central place in education policy. Governments recruit, train, rotate and evaluate them on the premise that the person who runs a school shapes how much its students learn, and a large empirical literature has sought to measure that influence. The answers vary widely. Estimates of the standard deviation of principal effects on achievement range from 0.28–0.40 student standard deviations in Canadian data (Dhuey and Smith 2014) to about 0.20 in Texas (Branch, Hanushek, and Rivkin 2012), and a synthesis of two decades of research places the typical value near 0.10 (Grissom, Egalite, and Lindsay 2021). The most recent evidence for the United States concludes that little of the variation in average test scores reflects persistent differences in principal effectiveness (Bartanen, Husain, and Liebowitz 2024). For Brazil, estimates based on principal fixed effects report standard deviations between 0.77 and 0.89 (Cauduro 2018), far above any of these benchmarks.

Much of this dispersion has an econometric origin. Principal effects are usually recovered as fixed effects estimated from a small number of school-years per principal. The variance of the estimated effects then adds sampling noise to the true variance, while the covariance between principal and school effects is biased downward when few principals move between schools (Andrews et al. 2008). Leave-out estimators correct both problems (Kline, Saggio, and Sølvesten 2020), but they require a persistent identifier for each principal and a connected network of moves across schools. Many education systems publish neither. In Brazil, the national assessment microdata anonymize school codes in every wave except 2011 and contain no principal identifier at all, and identified records can be consulted only in a secure room at the national education statistics agency, INEP.

This paper develops a design that does not require principal identifiers and applies it to the Brazilian public school system. Rather than exploiting the mobility of managers across schools, we exploit the discontinuity of management within schools. The 2011 wave of the national assessment (SAEB/Prova Brasil) is the only one that publishes true school codes, and its principal questionnaire records how long the current principal has led the school. Because the assessment is administered every November, the reported tenure band dates each school’s most recent succession relative to the 2005, 2007, 2009 and 2011 tests. For every combination of tenure band and two-year testing window, the probability that the school changed principal is therefore known by construction. Linking these probabilities to the school-level achievement panel for 2005–2023 and to the annual school census yields a nationwide panel of dated successions.

Identification rests on an event study in second moments. If school achievement is the sum of a school effect, the effect of the incumbent management, and noise, a succession adds to the variance of the two-year change in achievement a term equal to twice the variance of management effects. This excess variance should appear in exactly the windows, and in exactly the proportions,

implied by the transition probabilities. Group-by-window fixed effects absorb differences in baseline volatility between schools with and without recent successions, so the parameter is identified from the timing of successions within groups rather than from comparisons across them. A complementary covariance estimator isolates the persistent component of management effects and is immune to transitory disruption around the change. We validate both estimators with a Monte Carlo exercise that preserves the exact structure of the estimation sample, including its schools, groups, cells, cohort sizes and clusters.

The second part of the paper asks what happens when principals are replaced for political reasons. In Brazil's municipal school networks, many principals serve at the discretion of the mayor, whereas principals in state networks, which frequently operate schools in the same municipality, do not answer to local government. Mayoral elections held every four years therefore generate variation in principal turnover together with an internal control group. We study the 2008, 2012, 2016 and 2020 elections, first with triple-difference designs and then with a regression discontinuity that pools close elections across the four cycles.

The closest evidence on this question comes from Akhtari, Moreira & Trucco (2022). Using close mayoral elections in Brazil, they show that a change in the mayor's party brings new personnel into the municipal bureaucracy at both managerial and non-managerial levels, that test scores in schools controlled by the municipal government fall by 0.05–0.08 standard deviations, and that schools outside municipal control are unaffected. Our political analysis builds directly on their design, and part of it confirms their findings: party turnover replaces municipal principals even in close elections, and state schools in the same municipalities are unaffected. The paper differs from theirs in three questions it can answer. First, because the first part estimates the dispersion of principal effects, we can ask how much of the effect of political turnover could plausibly run through the principal's office, which designs based on turnover alone cannot determine because principals and teachers are replaced together. Second, because the rule by which each network selected its principals is observed before the elections we study, we can test their interpretation that turnover harms services where the bureaucracy is not shielded from politics. Third, because we compare municipal with state schools in the same municipality over four electoral cycles, we can separate the effect of turnover on the network the mayor controls from the deterioration that precedes it and that panel designs attribute to the change of government.

The first set of results concerns who changes principals and what a change does on average. Successions are selected on trajectories: schools that changed principals between 2009 and 2011 had grown 0.066 student standard deviations less than otherwise similar schools between 2005 and 2009, so that stable management is in part a consequence of school success. After the change, the relative decline stops, but there is no average short-run gain. Depending on whether one assumes parallel trends or selection on lagged outcomes, the short-run effect of a succession lies between

−0.031 and −0.002 SD in grade 5.

The second set of results concerns the dispersion of principal effects. The standard deviation of principal-by-school effects is 0.050 student SD in grade 5, with an upper bound of 0.068, and statistically zero in grade 9, with an upper bound of 0.041. The persistent-covariance estimator gives the same order of magnitude. Applied to the same data, the plug-in procedure used in much of the literature yields 0.176, and 0.344 when expressed in units of the between-school standard deviation, as is common practice. Plug-in variances, few observations per spell and standardization by the school mean are together sufficient to generate estimates of the size previously reported for Brazil. In simulations calibrated to our sample, the estimator is unbiased under the design assumptions, has correct coverage and 83% power at a true standard deviation of 0.05, and is biased mainly downward under plausible violations of those assumptions.

The third set of results shows that the rule by which principals are chosen shapes this dispersion. In grade 5, the variance of management effects is close to zero in networks that select principals through examinations or school elections and corresponds to a standard deviation of about 0.08 in networks that appoint them. The difference is statistically significant ($p = 0.021$, and $p = 0.008$ within municipal networks), survives controls for rurality, socioeconomic status and region interacted with the succession probability, and is not driven by any single state, although randomization inference is more conservative (one-sided $p = 0.053$). The first moment points in the same direction: among successions, competitively selected successors produce achievement 0.026 SD higher than appointed successors, with identical trends before the change.

The fourth set of results concerns political turnover. In close 2008 elections, which are the only cycle for which principal arrivals can be dated, a change in the mayor's party raises the municipal-minus-state difference in the share of recently arrived principals by 20 percentage points, and the increase is concentrated in the window containing the new mayor's inauguration and in networks where the post is filled by appointment. Triple-difference estimates suggest that turnover harms learning, but longer pre-periods show that party turnover follows a relative deterioration of municipal schools, as retrospective voting would imply. In the stacked regression discontinuity, where this deterioration disappears, turnover has no detectable effect on average achievement. The confidence interval excludes declines larger than 0.10 SD in grade 5 but contains the estimates of Akhtari, Moreira & Trucco (2022), so the two sets of results are statistically compatible even though our point estimate is smaller. The variance estimates show that principal replacement cannot account for most of an effect of that size. Given a first stage of 20 points, a decline of 0.05–0.08 SD would require politically induced successors to be worse than their predecessors by 0.11–0.39 student SD on average, depending on the units in which the decline is measured, which is at least 1.6 times the upper bound on the standard deviation of principal effects and at least 3.5 times the largest average change we measure at successions. Even if every induced successor were worse by the full

gap we estimate between appointed and competitively selected principals, the principal channel would account for between a tenth and a quarter of the decline. The average also conceals a pattern that supports the interpretation of Akhtari, Moreira & Trucco (2022) and that the variance results anticipate. Where principals were appointed before the election, turnover raises a summary index of harm to achievement and grade progression by 0.25 standard deviations, driven by grade 6–9 age-grade distortion rather than by test scores, while it leaves the index unchanged where principals are selected competitively. Teacher records show that the shock reaches beyond the principal’s office. In the same networks, turnover sharply lowers INEP’s indicator of teacher regularity, which measures how continuously a school’s teachers have taught there, whereas it leaves state schools in the same municipalities, the number of teachers and the mix of teaching contracts unchanged.

A final result concerns the management measures commonly used to study principals. A teacher-rated leadership index with a school-level reliability of 0.76 predicts achievement growth before the current principal arrived as strongly as growth after the arrival. In this setting, cross-sectional associations between management indices and achievement largely reflect the trajectories of the schools principals inherit.

Taken together, the results suggest that the identity of an individual principal accounts for little of the variation in school performance in Brazil, but that the institutions governing how principals are chosen matter in two ways. Discretionary appointment is associated with a wider dispersion of management quality, and it exposes schools to a political cycle that reshuffles their staff. These two findings are linked by a simple model of selection, developed in Section 3, in which appointment weighs political considerations that are unrelated to managerial quality.

The paper contributes to three literatures. The first estimates the contribution of managers to organizational performance, in schools (Branch et al. 2012; Coelli and Green 2012; Dhuey and Smith 2014; Bartanen et al. 2024) and in the public sector more broadly (Fenizia 2022). We propose an estimator that replaces the mobility of managers across units with dated discontinuities of management within units, which can be applied wherever surveys record tenure, and we show how much of the cross-study dispersion in estimated principal effects the choice of estimator alone can generate. The second literature studies patronage and political turnover in public bureaucracies (Colonnelli et al. 2020; Akhtari et al. 2022; Toral 2024a). Relative to Akhtari, Moreira & Trucco (2022), who establish that political turnover disrupts municipal school personnel and lowers test scores, we bound the share of this effect that can operate through principals, show that the replacement of principals and its costs are confined to networks where the post is discretionary, and show that the disruption extends to the continuity of the teaching staff. We also document a specific way in which panel designs can overstate these costs, because voters remove incumbents who preside over deteriorating services (Ferraz and Finan 2008). The third literature links management practices to outcomes (Bloom et al. 2015; Tavares 2015; Lemos et al. 2024). We show that, with the data

typically available to researchers, practice and leadership indices behave as much as outcomes of school success as inputs to it.

The questions we study are also of immediate policy relevance. Brazil’s national education plan (Law 13,005/2014, goal 19; [Brazil 2014](#)) calls for principal selection based on technical merit and community participation, and the law that made the federal equalization fund permanent (Fundeb, Law 14,113/2020, art. 14; [Brazil 2020](#)) conditions part of an outcome-based federal transfer, the VAAR component, on principals being selected by technical criteria of merit and performance or by the school community from previously approved candidates. Our estimates suggest that such rules are unlikely to raise average achievement substantially through the identity of the principals they select, because the dispersion of principal effects is small. They are better understood as a protection against the lower tail of management quality and against the disruption that follows changes of government.

The rest of the paper is organized as follows. Section 2 describes the institutional setting and Section 3 develops the conceptual framework. Section 4 presents the data and Section 5 the estimators. Section 6 reports the results on successions and on the variance of principal effects, and Section 7 the results on political turnover, principals and teachers. Section 8 discusses threats to validity, Section 9 interprets the findings and draws out their policy implications, and Section 10 concludes.

2 Institutional Background

2.1 Two public networks in each municipality

Brazilian public education in grades 1–9 is provided by two networks that often operate side by side within the same municipality. Municipal governments run most schools offering grades 1–5, while state governments run a larger share of schools offering grades 6–9, so that in most municipalities a state school and a municipal school serving similar students coexist. Each network sets its own rules for staffing, pay and the appointment of principals within a common national framework. Teachers are recruited by the network, either as civil servants through public examinations or on temporary contracts, and are assigned to schools by the education secretariat. Principals therefore have limited formal authority over hiring and dismissal. Their influence runs mainly through the organization of the school, pedagogical coordination, the management of teachers already in post and relations with families and the secretariat.

2.2 How principals are chosen

Networks differ sharply in how they fill the principal's post. In the 2011 SAEB principal questionnaire, 22.5% of public-school principals reported having been appointed by politicians, 13.3% through other forms of appointment and 11.8% through technical appointment. Another 20.5% had been elected by the school community, 13.6% selected through a combination of examination and election and 10.1% selected by examination alone, while 8.3% reported another category. The two networks differ systematically. Political appointment accounts for 30.5% of municipal principals and 9.7% of state principals, whereas examination and election combined account for 23.7% of state principals and 7.3% of municipal principals.

National legislation has pushed networks toward merit-based selection. Goal 19 of the 2014 national education plan calls for the appointment of principals on the basis of technical merit combined with community consultation. The 2020 law that made Fundeb permanent went further, conditioning part of the new outcome-based federal transfer (VAAR) on networks selecting principals by technical criteria of merit and performance, or by the school community among candidates previously approved on such criteria. Whether rules of this kind improve school management is the policy question to which Section 6.5 speaks.

2.3 Mayors and the school bureaucracy

Mayors are elected in October every four years and take office on January 1 of the following year. Where principals serve at the mayor's discretion, a change of government can trigger replacements early in the new term, as it does for other appointed posts in Brazilian municipalities (Colonnelli, Prem, and Teso 2020; Akhtari, Moreira, and Trucco 2022). Recent evidence shows that the disruption can begin even before the new government takes office, as outgoing administrations dismiss temporary staff after losing an election (Toral 2024a). State schools follow a different calendar. Governors are elected two years after mayors, and state principals do not respond to municipal elections, which makes state schools in the same municipality a natural comparison group for municipal schools.

Principal turnover is high throughout the system. In November 2011, 18.7% of public-school principals had led their school for less than one year, 14.4% for one to two years and 34.1% for two to five years, and only 21.2% had been in post for seven years or more (Table 1, Panel A). This rate of turnover is what makes the design in Section 5 feasible: a single cross-section of tenures dates a succession in most schools within a six-year window.

3 A Framework for Principal Selection and Turnover

This section makes explicit the object we estimate and the predictions that organize the empirical analysis. The framework is deliberately simple. Its purpose is to link the two halves of the paper, the dispersion of principal effects and the consequences of political turnover, through a single mechanism: the weight that the selecting authority places on managerial quality.

3.1 The object of interest

Let θ_{ks} denote the contribution of principal k to achievement in school s , measured relative to the school's long-run level. When principal k is replaced by principal k' , achievement changes by $\theta_{k's} - \theta_{ks}$. If the two effects are independent draws with variance σ_θ^2 , the expected squared change is $2\sigma_\theta^2$. The standard deviation σ_θ therefore governs the gains and losses that policies acting on the identity of principals, whether by replacing, selecting or rotating them, can produce. It is a variance of principal-by-school effects. It includes any component that depends on the match between a principal and a school and thus differs from the variance of transferable principal effects identified by designs based on mobility across schools (Kline, Saggio, and Sølvesten 2020). Because $\text{Var}(\theta_{ks})$ is the sum of the transferable and match components when they are independent, our estimates bound the transferable component from above.

3.2 Selection rules and the distribution of principal quality

Suppose that candidates for the post have managerial quality $\theta \sim N(\mu, \sigma_0^2)$ and that the selecting authority ranks them by an index $I = \theta + \nu$, where ν collects everything other than quality that drives the choice. Under competitive selection, ν is the error with which an examination or a school election measures quality. Under discretionary appointment, it also includes political loyalty, connections and other attributes valued by the appointing authority. Let $r^2 = \sigma_0^2 / (\sigma_0^2 + \sigma_\nu^2)$ be the share of the variance of the index that reflects quality. If the authority appoints candidates whose index exceeds a threshold that admits a share p of the pool, standard results on selection on a correlated variable give

$$\mathbb{E}[\theta \mid \text{selected}] = \mu + r \sigma_0 \lambda(c), \quad \text{Var}(\theta \mid \text{selected}) = \sigma_0^2 [1 - r^2 \delta(c)], \quad (1)$$

where $c = \Phi^{-1}(1 - p)$, $\lambda(c) = \phi(c) / [1 - \Phi(c)]$ is the inverse Mills ratio and $\delta(c) = \lambda(c)[\lambda(c) - c]$ lies between zero and one. Selection on an index that is informative about quality raises the mean and compresses the dispersion of the principals selected, and both effects grow with r^2 . When appointment gives substantial weight to attributes unrelated to quality, r^2 is small, and appointed principals resemble random draws from the candidate pool.

Two predictions follow. The first (P1) is that the variance of principal effects should be larger in networks that appoint principals than in networks that select them competitively. The second (P2) is that competitively selected successors should, on average, outperform appointed successors. Equation (1) also disciplines the interpretation of magnitudes. Taken literally, a ratio of standard deviations between competitive and appointment networks of 0.017 to 0.078, which is what we estimate in grade 5, would require $r^2\delta(c) \approx 0.95$, that is, both a highly informative selection procedure and intense screening. We therefore read the model as a guide to the direction of the differences rather than to their size. Other features of appointment networks, such as the discretion principals enjoy over staff once appointed, may amplify the dispersion of their effects.

3.3 Political turnover

Part of what makes an appointed principal valuable to a mayor is loyalty to the governing party. When the party changes, that value falls for the incoming administration, and the new mayor has an incentive to replace principals who serve at the mayor’s discretion. Where the post is filled by examination or election, the mayor cannot easily do so. The third prediction (P3) is therefore that party turnover increases principal replacement only in networks where the post is discretionary.

The consequences of the induced replacements for average outcomes can be bounded. If party turnover replaces a share f of a network’s principals, the direct effect on the network mean is $f \times (\mathbb{E}[\theta_{\text{new}}] - \mathbb{E}[\theta_{\text{old}}])$. With the first stage we estimate ($f \approx 0.2$) and a gap between successor types no larger than the one we observe (0.026 SD), the direct effect is of the order of 0.005 SD, well below what any design based on public data can detect. The fourth prediction (P4) is that close-election estimates of the effect of turnover on average achievement should be close to zero. The fifth (P5) follows from the same logic. If turnover has detectable effects on school outcomes, they must operate through channels other than the average quality of new principals, such as the disruption of the transition itself or the replacement of other staff who also serve at the mayor’s discretion, and they should therefore appear where the post is discretionary.

Finally, electoral outcomes are not independent of school performance. Voters observe local public services and remove incumbents who preside over their deterioration (Ferraz and Finan 2008). If municipal schools decline before an election, party turnover will be correlated with pre-existing trends, and panel designs will attribute that decline to the change of government. Close elections break this correlation, which is why we rely on them for causal inference.

The empirical analysis follows the order of these predictions. Section 6.5 tests P1 and P2, and Section 7 tests P3 to P5 and documents the retrospective-voting problem in panel designs.

4 Data

4.1 Principal and teacher questionnaires

The 2011 wave of SAEB covers 58,960 schools with principal questionnaires and 316,668 teacher respondents. From the principal questionnaire we use tenure at the current school, reported in eight bands from less than one year to more than twenty, years in principal positions overall, the mode of access to the post, and thirteen items on management practices. From the teacher questionnaire we build a school-level leadership index from eight Likert items. The index has a Cronbach's α of 0.935 at the teacher level and an intraclass correlation of 0.443, which with 4.06 teachers per school on average implies a reliability of 0.764 for school means. SAEB anonymizes school identifiers in every other wave, and INEP ensures that anonymized identifiers do not overlap across waves. No public file contains a persistent principal identifier. We use only identifiers published by INEP and do not attempt to re-identify schools or principals in anonymized waves.

4.2 Achievement and the school census

INEP publishes school-level mean proficiency in mathematics and Portuguese for grades 5 and 9 every two years from 2005 to 2023, with true school codes, together with the pass-rate indicator used to compute the national quality index (IDEB). We convert scale scores into student standard deviations by dividing by the 2011 individual-level standard deviation in public schools, which is 47.4 in mathematics and 45.7 in Portuguese in grade 5, and 47.0 and 46.7 in grade 9. This choice matters for comparisons with the literature. The between-school standard deviation of mean scores is 0.43–0.55 of the between-student standard deviation, so effects expressed in units of the school-mean distribution, as is common, are roughly twice as large as the same effects expressed in student units. The annual School Census, with a reference date in May, provides each school's administrative dependency, location, enrollment and infrastructure. Enrollment in the tested stage divided by the number of grades approximates the size of the tested cohort, which governs the sampling noise of school means.

4.3 School flow

We assembled an annual school panel of pass, repetition and dropout rates, measured at the end of the school year in December, and of age-grade distortion rates, measured in May, for grades 1–5 and 6–9 from INEP's official spreadsheets for 2007–2024. The panel contains about 2.5 million school-years. The files use at least three coding schemes across years, and in the 2018 release the variable codes are shifted relative to their labels, so we match columns by header label and audit

the choice year by year. At the seams between formats, school-to-school correlations are the same as those between adjacent years; for example, the correlation of the grade 6–9 pass rate between 2017 and 2018 is 0.683, against 0.676 and 0.673 for the neighboring pairs.

4.4 Teaching staff

The person-level teacher files are no longer part of the public microdata, but two INEP products measure the teaching staff without individual identifiers. The teacher regularity indicator (IRD), published by school from 2016, scores how continuously each of the school’s teachers has taught there over the preceding five years on a scale from 0 to 5 and averages across teachers, so that the arrival of new teachers lowers it. The statistical synopses of the School Census for 2012–2022 report, for each municipality and network, the number of teachers in grades 1–5 and 6–9 by employment status: permanent civil servant, temporary contract, outsourced or CLT (private-sector labor contract). Because SAEB also anonymizes municipality codes from 2013 onward, principal turnover cannot be measured by municipality after 2011 with public data, and these teacher measures are the only public window on the staff of municipal schools in later electoral cycles.

4.5 Municipal elections

Official results by municipality for 2004–2020 come from the electoral database of the Center for Metropolitan Studies (CEM, University of São Paulo; [Centro de Estudos da Metrópole n.d.](#)), which compiles Superior Electoral Court results with IBGE municipality codes. For each election we identify the winning party, whether a runoff took place and the incumbent party’s first-round margin over its strongest rival. The ruling party changed in 63.3% of municipalities in 2008, 69.6% in 2012, 75.5% in 2016 and 66.8% in 2020, and the incumbent party ran in 53–60% of municipalities, depending on the cycle.

5 Empirical Strategy

5.1 A statistical model of school achievement

For school s in test wave t , we write

$$y_{st} = \phi_s + \lambda_{c(s)t} + \theta_{k(s,t)} + \varepsilon_{st}, \quad \text{Var}(\varepsilon_{st} \mid n_{st}) = \sigma_\eta^2 + \frac{\sigma_a^2}{n_{st}}, \quad (2)$$

where ϕ_s is a school effect, $\lambda_{c(s)t}$ is a shock common to the school’s cell (state $\times$ network $\times$ location), $\theta_{k(s,t)}$ is the effect of the management spell k in place at t , and ε_{st} combines transitory

school shocks with variance σ_η^2 and the sampling error of a mean over n_{st} students. The parameter of interest is $\sigma_\theta^2 = \text{Var}(\theta_k)$, the variance of principal-by-school effects defined in Section 3.1.

5.2 Dating successions

Because tests are administered each November, the tenure band reported in 2011 implies a window for the arrival of the current principal (Table 1, Panel B). Schools fall into four groups. In group T, the principal had been in post for less than two years and therefore arrived between November 2009 and November 2011. In group C, tenure of two to five years implies an arrival between November 2006 and November 2009, so that the succession falls in the 2005–07 testing window with probability 1/3 and in the 2007–09 window with probability 2/3. In group D, tenure of five to seven years implies an arrival between November 2004 and November 2006, so that the succession falls before 2005 or in 2005–07 with probability 1/2 each. In group S, tenure of seven years or more implies that the same principal has led the school since before the 2005 test.

Under uniform arrivals within tenure bands, the probability $\pi_{g,w}$ that a school in group g experienced a succession in window w is known. It equals one for group T in 2009–11, 2/3 for group C in 2007–09 and 1/2 for group D in 2005–07, and zero for group S in every window, for group C in 2009–11 and for group D in 2007–09 and 2009–11. For windows before the arrival of the T principal, we do not know whether the predecessor had also arrived recently. We treat that probability, q , as a sensitivity parameter, with a baseline value of 0.33 and a range of 0.20–0.50, and we report estimates that exclude group T altogether.

Table 1—Principal tenure in 2011 and the dating of successions

<i>Panel A. Tenure at the current school, public-school principals, SAEB 2011 (N = 55,372)</i>								
Tenure	< 1 yr	1–2	2–5	5–7	7–10	10–15	15–20	> 20
Share (%)	18.7	14.4	34.1	11.6	9.9	7.3	2.6	1.4
<i>Panel B. Succession groups (grade-5 schools with scores in 2009 and 2011)</i>								
Group	Tenure in 2011	Arrival of current principal	Succession window (probability)		Schools			
T	< 2 years	Nov 2009 – Nov 2011	2009–11 (1)		11,469			
C	2–5 years	Nov 2006 – Nov 2009	2005–07 (1/3) or 2007–09 (2/3)		12,007			
D	5–7 years	Nov 2004 – Nov 2006	before 2005 (1/2) or 2005–07 (1/2)		4,152			
S	≥ 7 years	before Nov 2004	none in 2005–11		7,447			

5.3 Excess variance in the succession window

Let $\Delta y_{s,w}$ denote the two-year change in achievement in window w , and $u_{s,w}$ its residual after removing cell $\times$ window and group $\times$ window means. Under model (2),

$$\mathbb{E}[u_{s,w}^2] = a_g + b_{c,w} + 2\sigma_\theta^2 \pi_{g,w} + \kappa \left(\frac{1}{n_w} + \frac{1}{n_{w-1}} \right), \quad (3)$$

where the group effects a_g absorb differences in baseline volatility and in the level of selection across groups, and $b_{c,w}$ are cell $\times$ window effects. We estimate equation (3) by least squares, clustering standard errors by municipality. Identification comes from variation in $\pi_{g,w}$ within groups across windows: group C should be more volatile in 2007–09 than in 2009–11, group D more volatile in 2005–07 than afterwards, and group S equally volatile throughout. The identifying assumption is that, conditional on group and window, transitory shocks are not more volatile in the window in which a succession occurred than in other windows. This assumption would fail if successions were triggered by unusually large shocks, a possibility we examine in Section 5.5 and 8. The coefficient κ provides a built-in check on the specification. Under the model it should approximate the within-school variance of student scores, which is close to one in squared student-SD units.

5.4 Excess persistent covariance

A succession could raise the variance of changes through a transitory disruption at the time of the change rather than through persistent differences in management. To isolate the persistent component, we use groups whose new principal is still in post at the following measurement. For group C, with base year 2007, let $D_1 = y_{2009} - y_{2007}$ and $D_2 = y_{2011} - y_{2007}$. Then

$$\text{Cov}(D_1, D_2) = 2\sigma_\theta^2 \pi + \text{Var}(\varepsilon_{2007}), \quad (4)$$

so that, relative to group S, the excess covariance identifies $2\sigma_\theta^2 \pi$. Transitory shocks in 2009 and 2011 do not enter this expression. The analogue for group D uses 2005 as the base year and measurements in 2007 and 2009. For group T, with base year 2009 and measurements in 2011 and 2013, the estimator is a lower bound, because the new principal may have left by 2013.

5.5 Validation by calibrated simulation

Because both estimators rely on assumptions about the timing of arrivals and the behavior of shocks, we validate them with a plasmode Monte Carlo exercise that keeps the exact estimation sample: the 37,264 grade-5 schools, their groups, cells, windows, cohort sizes and municipalities. We simulate $y_{st} = \theta_{k(s,t)} + \eta_{st} + e_{st}$, drawing a new management effect at each succes-

sion. The noise process is calibrated on stable schools, where the variance of two-year changes is $0.0487 + 0.974 (1/n_a + 1/n_b)$ in grade 5 and $0.0378 + 1.150 (1/n_a + 1/n_b)$ in grade 9. This implies a school-year shock with a standard deviation of 0.156 and sampling error equal to that of a mean of n students with unit variance. The simulated data are passed through the same code as the actual estimates. We use 1,000 replications for each value of $\sigma_\theta \in \{0, 0.025, 0.05, 0.075, 0.10, 0.15\}$ under the design assumptions and 400 replications under each of seven violations of those assumptions, which include misreported tenure, arrivals bunched after the 2009 inauguration, effects that build up with tenure and successions triggered by negative shocks.

5.6 Triple differences

To study political turnover, we first estimate for each election cycle $e \in \{2008, 2012, 2016, 2020\}$

$$y_{smt} = \sum_{\tau \neq -1} \beta_\tau \text{Alt}_{m,e} \times \text{Mun}_s \times \mathbf{1}[t = e + \tau] + \alpha_s + \gamma_{mt} + \delta_{r(s)n(s)t} + \mu_{n(s)l(s)t} + \varepsilon_{smt}, \quad (5)$$

where $\text{Alt}_{m,e}$ indicates that the party winning the mayoral election in municipality m differs from the previous winner and Mun_s indicates a municipal school. Municipality $\times$ year effects γ_{mt} absorb all shocks common to the two networks in the municipality, including the local economic and political conditions that accompany a change of government, and δ and μ are state $\times$ network $\times$ year and network $\times$ location $\times$ year effects. The reference period is the year before the election. The identifying assumption is that, absent party turnover, the gap between municipal and state schools would have evolved in the same way in municipalities with and without turnover. We also stack cycles in event time with cycle-specific fixed effects and estimate post-election effects as deviations from the linear pre-election trend of the treated group, in the spirit of sensitivity analyses that ask how much of an estimated effect a pre-existing divergence could explain (Rambachan and Roth 2023).

5.7 Stacked regression discontinuity

Because the parallel-trends assumption turns out to fail, our preferred design is a regression discontinuity in close elections. The unit of observation is the municipality $\times$ cycle, and the outcome is the internal contrast used throughout,

$$\Delta Y_{mc} = \overline{\Delta y_{mc}}^{\text{mun}} - \overline{\Delta y_{mc}}^{\text{state}}, \quad (6)$$

the mean change in municipal schools minus the mean change in state schools in municipality m around election c . For achievement, the change runs from the last test before the election to the first test after it. For school flow, it runs from the year before the election to the mean of the first three years of the new term. Each school’s network is measured in the year before the election. The running variable X_{mc} is the incumbent party’s first-round margin of defeat, and the estimand is

$$\tau_{RD} = \lim_{x \downarrow 0} \mathbb{E}[\Delta Y_{mc} \mid X_{mc} = x] - \lim_{x \uparrow 0} \mathbb{E}[\Delta Y_{mc} \mid X_{mc} = x]. \quad (7)$$

We restrict attention to municipalities where the incumbent party ran and no runoff took place, which yields 12,527 municipality-cycles. In this sample, party turnover coincides exactly with a positive margin of defeat in all four cycles, so the design is sharp. We estimate local linear regressions with a triangular kernel and MSE-optimal bandwidths, report bias-corrected robust confidence intervals (Calonico, Cattaneo, and Titiunik 2014), include cycle indicators as covariates and cluster errors by municipality. We test the continuity of the density of the running variable (Cattaneo, Jansson, and Ma 2020) and the balance of pre-election levels and pre-election changes. Differencing the outcome within municipality and across networks removes the component of any electoral shock that affects both networks alike, so the design identifies the effect of turnover on the network the mayor controls relative to the one the mayor does not.

6 Successions and the Variance of Principal Effects

6.1 Which schools change principals

Successions are selected on trajectories rather than on levels (Table 2). Relative to stable schools in the same state, network and location, schools with a 2009–11 succession have only slightly lower socioeconomic status (-0.025 SD) and 2005 scores (-0.035 student SD). They had grown markedly less, however: 0.066 SD less between 2005 and 2009 (s.e. 0.008) and 0.049 SD less between 2007 and 2009 (s.e. 0.008). The gradient is monotonic in the time elapsed since the current principal’s arrival. Mean growth between 2005 and 2009 is 0.526 in stable schools, 0.494 in schools whose principal arrived in 2004–06, 0.466 for arrivals in 2006–09 and 0.443 in schools with a succession in 2009–11, a pattern that is difficult to reconcile with random turnover and easy to reconcile with principals leaving, or being removed from, schools that are falling behind.

The principals themselves differ as well. Principals of succession schools are 10.7 percentage points more likely to have been appointed and 17.4 points more likely to have led another school before, and 8.0 points less likely to hold a graduate degree. Teacher-rated leadership is 0.213 SD lower and principals’ self-reported practices are 0.313 SD lower, differences that partly reflect the short tenure of new principals. These contrasts are a reminder that cross-sectional comparisons

between schools with new and established principals confound the effect of management with the circumstances that led to the change.

Table 2—Balance: successions versus stable management (grade 5)

Variable	Stable ≥ 7 yrs	Arrival 2004–06	Arrival 2006–09	Succession 2009–11	Difference, succession – stable (s.e.)
School SES (z)	–0.023	–0.052	–0.088	–0.076	–0.025 (0.003)
Math 2005 (student SD)	3.844	3.791	3.773	3.769	–0.035 (0.008)
Math 2009 (student SD)	4.311	4.224	4.165	4.146	–0.101 (0.008)
Δ Math 2005→09	0.526	0.494	0.466	0.443	–0.066 (0.008)
Δ Math 2007→09	0.303	0.283	0.264	0.241	–0.049 (0.008)
Female principal	0.877	0.868	0.845	0.823	–0.045 (0.006)
Graduate degree	0.794	0.797	0.757	0.714	–0.080 (0.008)
Any appointment	0.382	0.438	0.563	0.547	+0.107 (0.018)
Previously led another school	0.174	0.270	0.239	0.325	+0.174 (0.008)
Teacher-rated leadership (z)	0.165	0.066	–0.012	–0.090	–0.213 (0.016)
Self-reported practices (z)	0.247	0.160	0.038	–0.146	–0.313 (0.017)
Schools	7,447	4,152	12,007	11,469	

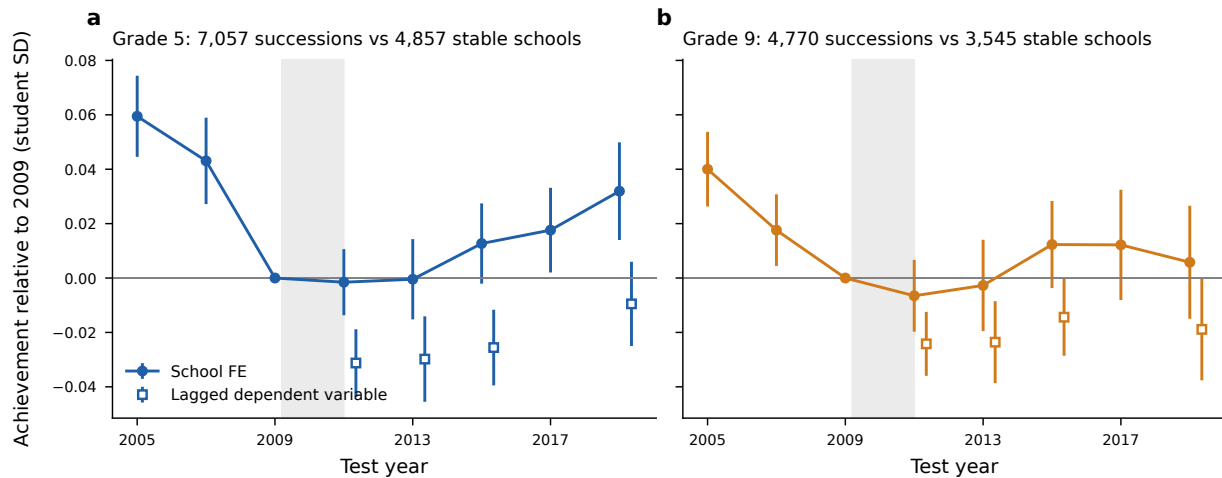
Notes: Differences estimated with state × network × location fixed effects; standard errors clustered by municipality. Scores in student-SD units (scale score divided by the 2011 individual SD).

6.2 The average effect of a succession

Figure 1 displays the event-study signature of this endogenous turnover. Relative to 2009, grade-5 achievement in succession schools was 0.059 SD above that of stable schools in 2005 and 0.043 SD above it in 2007; the schools that later changed principals had been falling behind. After the succession the decline stops (2011: –0.002, s.e. 0.006; 2013: –0.000, s.e. 0.008), and a gradual recovery follows from 2015 onward (+0.013 in 2015 and +0.032 in 2019).

Because the pre-succession decline is informative about the counterfactual, no single estimator is credible on its own. We therefore bracket the average effect between a fixed-effects estimator, which assumes parallel trends and attributes the pre-existing decline to school effects, and a lagged-dependent-variable estimator, which assumes selection on past outcomes. Because each estimator is biased in a known direction when the other assumption holds, the two bracket the effect (Angrist and Pischke 2009). The short-run effect lies between –0.031 and –0.002 SD in grade 5 and between –0.024 and –0.007 SD in grade 9 (Table 3). A change of principal therefore produces no average short-run gain and probably a small loss, similar in magnitude to U.S. estimates of the disruption caused by principal turnover (Bartanen, Grissom, and Rogers 2019). With minimum detectable effects of 0.017 SD in grade 5 and 0.019 SD in grade 9 at 80% power, the design is precise enough

Figure 1—Achievement around a change of principal



Notes: Difference in mean achievement (average of mathematics and Portuguese, student SD) between schools with a succession between November 2009 and November 2011 and schools with the same management since before 2005. School and state $\times$ network $\times$ location $\times$ year fixed effects; reference year 2009; 95% confidence intervals with errors clustered by municipality. Hollow squares: lagged-dependent-variable estimator conditioning on 2005, 2007 and 2009 scores. The shaded band marks the succession window.

to rule out the large average gains that the case for replacing principals of struggling schools would require.

Table 3—Bracketing the average effect of a succession (student SD)

Grade	Horizon	Fixed effects	Lagged dependent variable	Linear pre-trend extrapolation, M = 0.5	M = 1
5	2011	−0.002 (0.006)	−0.031 (0.006)	0.013	0.028
5	2013	0.000 (0.008)	−0.030 (0.008)	0.030	0.061
5	2015	0.013 (0.008)	−0.026 (0.007)	0.056	0.098
9	2011	−0.007 (0.007)	−0.024 (0.006)	0.003	0.013
9	2013	−0.002 (0.009)	−0.024 (0.008)	0.017	0.037
9	2015	0.013 (0.008)	−0.014 (0.007)	0.043	0.073

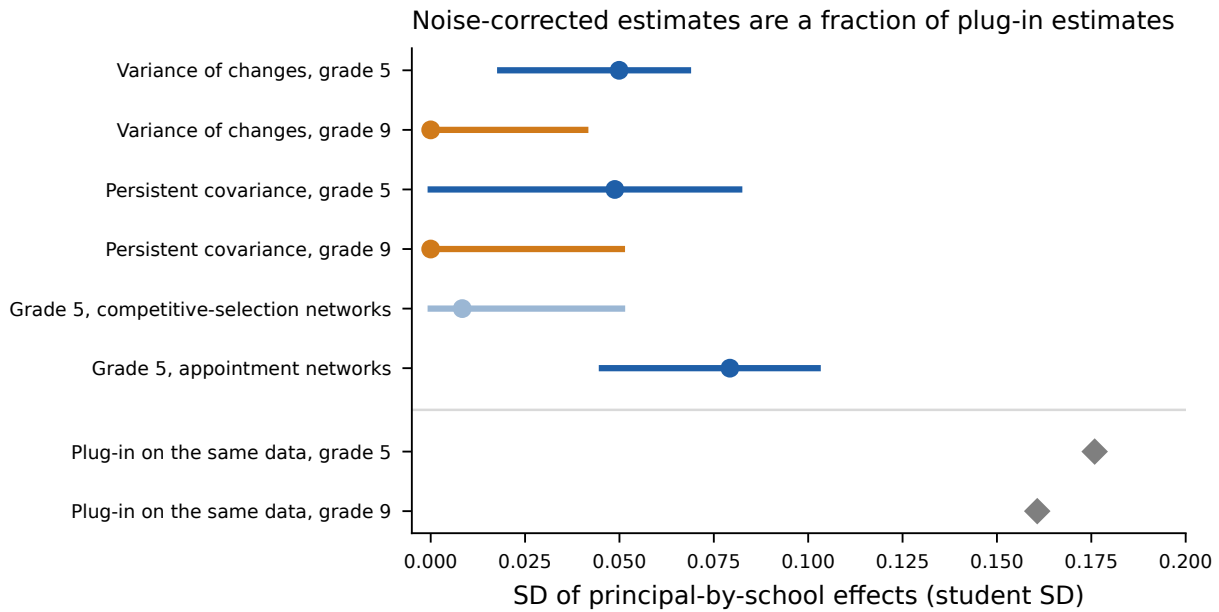
Notes: 7,057 succession and 4,857 stable schools in grade 5; 4,770 and 3,545 in grade 9. The pre-succession slope of group T is −0.030 (grade 5) and −0.020 (grade 9) SD per wave. Columns M = 0.5 and M = 1 extrapolate a fraction M of that slope (Rambachan and Roth 2023); full extrapolation is plausible only if the pre-succession decline would have continued indefinitely.

6.3 The variance of principal-by-school effects

The central estimates concern the second moment. The residual variance of two-year changes is similar across groups and barely rises in the windows that contain successions. In grade 5, the mean

squared residual of group T is 0.1047 in 2009–11 against 0.1038 in 2007–09, and that of group C is 0.1169 in 2007–09 against 0.1078 in 2009–11 (Appendix Table A1). The noise coefficient in equation (3) is $\kappa = 1.16$ (s.e. 0.05), close to its theoretical value of about one, which indicates that the specification captures the sampling component of the variance correctly.

Figure 2—Estimates of the standard deviation of principal-by-school effects



Notes: Student-SD units; 95% intervals from the interval for $2\sigma_\theta^2$, truncated at zero. “Variance of changes”: Section 5.3 with groups T, C, D and S ($q = 0.33$). “Persistent covariance”: Section 5.4, group C. Network rows: separate estimation by selection regime. Plug-in: standard deviation of differences between spell means in succession schools divided by $\sqrt{2}$, without noise correction.

Table 4 reports the estimates. The baseline standard deviation of principal-by-school effects is 0.050 student SD in grade 5, with a 95% upper bound of 0.068, and zero in grade 9, with an upper bound of 0.041. The estimates are stable across values of q and across subjects. Excluding group T, whose contribution depends on q , raises the grade-5 estimate to 0.075 with a wider interval. The persistent-covariance estimator, which is immune to transitory disruption at the time of the change, gives the same order of magnitude: 0.049 (upper bound 0.082) from group C and 0.038 (upper bound 0.087) from group D in grade 5. Across all specifications, the standard deviation never exceeds 0.11 student SD, and the most likely values lie between 0.00 and 0.08. In grade 9, where the covariance estimator gives 0.000 for group C and 0.066 for group D, the evidence is consistent with small effects but less precise.

How much confidence these magnitudes deserve depends on how the estimator behaves when its assumptions hold and when they fail. Figure 3 summarizes the Monte Carlo evidence. Under the design assumptions, the median estimate for a true σ_θ of 0.05 is 0.0497 in both grades, and

Table 4—Standard deviation of principal-by-school effects (student SD)

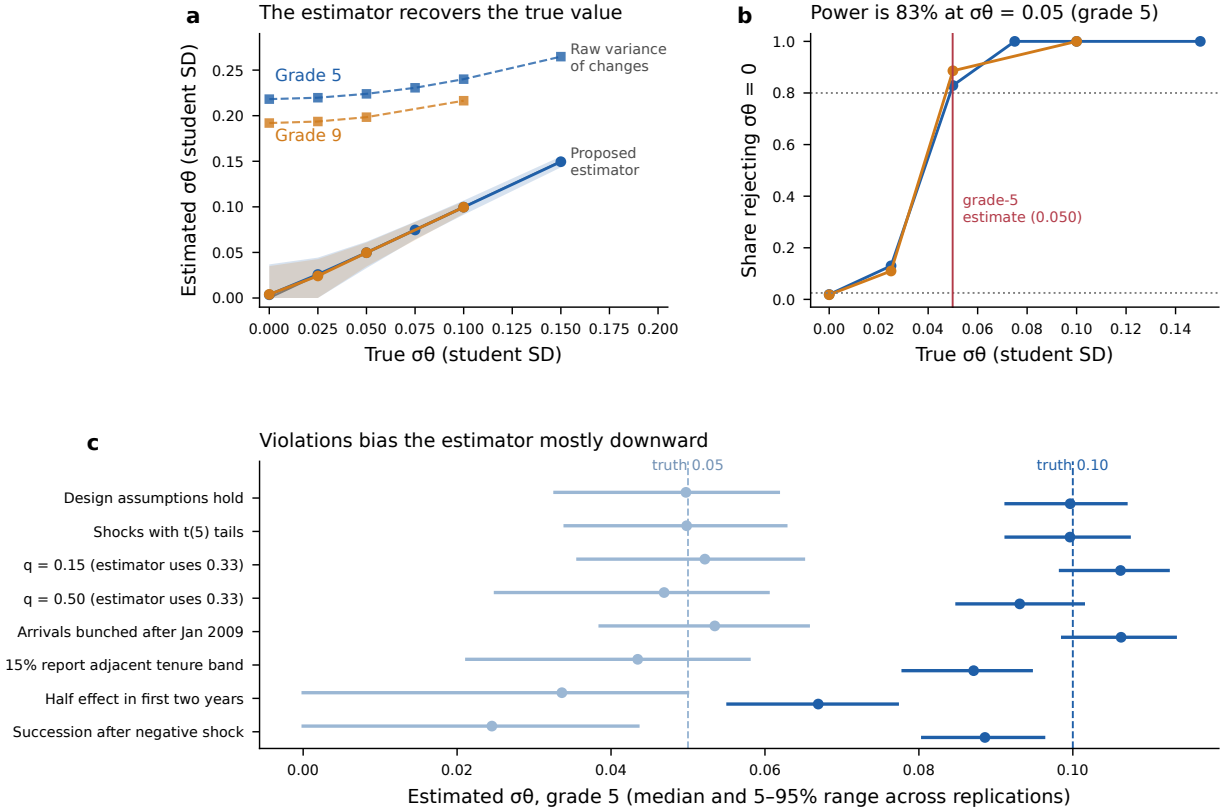
Grade	Outcome	Estimator / specification	$2\sigma_\theta^2$ (s.e.)	σ_θ	95% upper bound	Schools
5	Mean	Variance of changes, C, D, S (π known)	0.0111 (0.0046)	0.075	0.100	24,891
5	Mean	+ T, $q = 0.25$	0.0046 (0.0021)	0.048	0.066	37,264
5	Mean	+ T, $q = 0.33$ (baseline)	0.0050 (0.0022)	0.050	0.068	37,264
5	Mean	+ T, $q = 0.50$	0.0059 (0.0025)	0.055	0.074	37,264
5	Mathematics	+ T, $q = 0.33$	0.0043 (0.0028)	0.046	0.070	37,264
5	Portuguese	+ T, $q = 0.33$	0.0058 (0.0021)	0.054	0.070	37,264
9	Mean	Variance of changes, C, D, S (π known)	−0.0012 (0.0041)	0.000	0.059	19,207
9	Mean	+ T, $q = 0.33$ (baseline)	−0.0008 (0.0021)	0.000	0.041	28,567
5	Mean	Persistent covariance, group C	0.0032 (0.0029)	0.049	0.082	—
5	Mean	Persistent covariance, group D	0.0014 (0.0031)	0.038	0.087	—
9	Mean	Persistent covariance, group C	−0.0004 (0.0019)	0.000	0.051	—
9	Mean	Persistent covariance, group D	0.0043 (0.0035)	0.066	0.106	—

Notes: Cell $\times$ window and group fixed effects; errors clustered by municipality. For the covariance estimator the $2\sigma_\theta^2$ column reports the excess covariance relative to group S, which equals $2\sigma_\theta^2\pi$ ($\pi = 2/3$ for C, $1/2$ for D; Section 5.4), and σ_θ is recovered accordingly. The minimum detectable σ_θ (one-sided 5%, 80% power) is 0.052 in grade 5 and 0.051 in grade 9. The lower bound from group T is zero in grade 5 and 0.030 in grade 9.

the coverage of the 95% interval for $2\sigma_\theta^2$ lies between 93.6% and 96.2% at every value of σ_θ . The clustered standard error matches the standard deviation of the estimates across replications (0.0017 against 0.0017). With $\sigma_\theta = 0$, the one-sided rejection rate is 1.9% against a nominal 2.5%, and power to reject $\sigma_\theta = 0$ is 83% at $\sigma_\theta = 0.05$ in grade 5 (89% in grade 9) but only 13% at $\sigma_\theta = 0.025$. The grade-9 zero is therefore informative about effects of the size found in grade 5, but not about effects half as large.

The violations are more revealing. Only two bias the estimator upward, and only slightly: fewer prior successions in group T than assumed ($q = 0.15$) and arrivals of group C bunched after the January 2009 inauguration (probability 0.90 instead of $2/3$), which yield median estimates of 0.052 and 0.053 for a true value of 0.05. The others bias it downward. When 15% of principals report an adjacent tenure band, the median is 0.043. When principal effects are halved during the first two years of a spell, it is 0.034. When successions are triggered by a negative shock, in an extreme case in which the implied pre-succession decline is about 2.5 times the observed one, it is 0.025. Shocks with $t(5)$ tails change nothing. Inverting these results, the true σ_θ consistent with our grade-5 estimate lies between about 0.047 and 0.075.

Figure 3—Monte Carlo validation on the actual estimation sample



Notes: (a) Median (line) and 5–95% range (band) of the estimated σ_θ against its true value, with the raw variance of changes in the succession window (dashed), which attributes all variation to principals. (b) Share of replications rejecting $\sigma_\theta = 0$ (one-sided, 2.5%); dotted lines at 0.025 and 0.80. (c) Estimated σ_θ under violations of the design assumptions, for true values of 0.05 and 0.10. 1,000 replications per value under the design assumptions and 400 under each violation.

6.4 Why plug-in estimates are large

Applying the conventional plug-in procedure to the same data reproduces the large estimates of the literature, which allows us to trace the gap to specific steps. We compute spell effects as mean residuals, without any correction for noise, for the 6,974 grade-5 schools with a succession (Appendix Table A2). The standard deviation of the difference between the means of the two spells is 0.249 student SD, while the sampling noise expected in that difference, calibrated on stable schools, is 0.248. Once noise is netted out, a variance of 0.020 remains, which implies $\sigma_\theta = 0.014$. The plug-in estimate, by contrast, is 0.176 in student-SD units and 0.344 in units of the between-school standard deviation (0.365 in grade 9). The Monte Carlo makes the same point from a different angle: the raw variance of changes implies 0.22 when the true σ_θ is zero and only 0.26 when it is 0.15, so it is almost uninformative about the object it purports to measure. Plug-in variances, few observations per spell and standardization by the school mean are together sufficient to generate estimates of the

order of 0.8, which is what fixed-effects estimates for Brazil report (Cauduro 2018).

6.5 Selection rules and the dispersion of principal effects

We now turn to predictions P1 and P2. We classify each network, defined as a municipality $\times$ dependency pair, by the leave-one-out share of its principals who report having been appointed. Networks with a share below one third are classified as competitive and those with a share above two thirds as appointment networks. The distribution of this share is strongly bimodal: 80.2% of state schools belong to competitive networks, and 53.4% of municipal schools to appointment networks.

Table 5—Selection regime and principal effects

<i>Panel A. σ_θ by regime, separate estimation (student SD; 95% upper bound in parentheses)</i>					
Grade	Competitive networks		Appointment networks		
5	0.008 (0.051)		0.079 (0.103)		
9	0.017 (0.058)		0.000 (0.062)		
<i>Panel B. Pooled estimation with a $\pi \times$ regime interaction</i>					
Grade	Sample	Competitive	Appointment	Difference in $2\sigma_\theta^2$ (s.e.)	p
5	All networks	0.017	0.078	0.0117 (0.0050)	0.021
5	Municipal networks only	0.000	0.079	0.0153 (0.0057)	0.008
5	Excluding the Northeast	0.000	0.087	0.0157 (0.0066)	0.018
5	Municipal, excluding the Northeast	0.000	0.087	0.0194 (0.0077)	0.013
5	Cohorts of at least 20 students	0.012	0.075	0.0110 (0.0050)	0.029
5	$q = 0.50$	0.018	0.084	0.0136 (0.0058)	0.018
5	$q = 0.20$	0.017	0.074	0.0104 (0.0046)	0.024
9	All networks	0.012	0.000	−0.0011 (0.0054)	0.843
<i>Panel C. Average effect by type of successor (triple difference, $T \times$ post $\times$ Z)</i>					
Grade	Contrast Z	$T \times$ post $\times$ Z (s.e.; p)		Pre-trend $T \times$ pre $\times$ Z (s.e.; p)	
5	Successor selected or elected vs appointed	0.026 (0.009; 0.003)		−0.004 (0.009; 0.68)	
5	Competitive vs appointment network	0.022 (0.014; 0.11)		0.005 (0.014; 0.73)	
9	Successor selected or elected vs appointed	0.002 (0.009; 0.86)		0.004 (0.009; 0.65)	
9	Competitive vs appointment network	0.024 (0.016; 0.14)		−0.008 (0.014; 0.60)	

Notes: Panel B shares cell $\times$ window effects across regimes, so levels differ slightly from Panel A. For Portuguese alone, p-values in Panel B range from 0.008 to 0.032; for mathematics alone, from 0.076 to 0.114.

In grade 5, the variance of principal effects is concentrated in appointment networks, as P1 predicts (Table 5). Estimated separately, the standard deviation is 0.008 in competitive networks and 0.079 in appointment networks (Panel A). In a pooled regression that interacts the succession probability with the regime, the difference in $2\sigma_\theta^2$ is 0.0117 (s.e. 0.0050; $p = 0.021$), and it is larger

within municipal networks alone (0.0153, s.e. 0.0057; $p = 0.008$), a comparison that holds constant the level of government (Panel B). The difference also survives the exclusion of the Northeast, where appointment is more frequent and socioeconomic status lower, the combination of both restrictions, a minimum cohort size of 20 students and alternative values of q . For Portuguese alone, the p-values across these specifications range from 0.008 to 0.032, and for mathematics alone from 0.076 to 0.114. Grade 9, which is predominantly served by state networks and therefore offers little variation in regime, shows no difference.

The first moment points in the direction predicted by P2. Among successions, successors chosen by competitive selection or election produce grade-5 achievement 0.026 SD higher (s.e. 0.009; $p = 0.003$) than appointed successors, with no difference in trends before the change (-0.004 , s.e. 0.009; Panel C). The comparison between regimes, rather than between successor types, gives a similar point estimate (0.022) that is less precise.

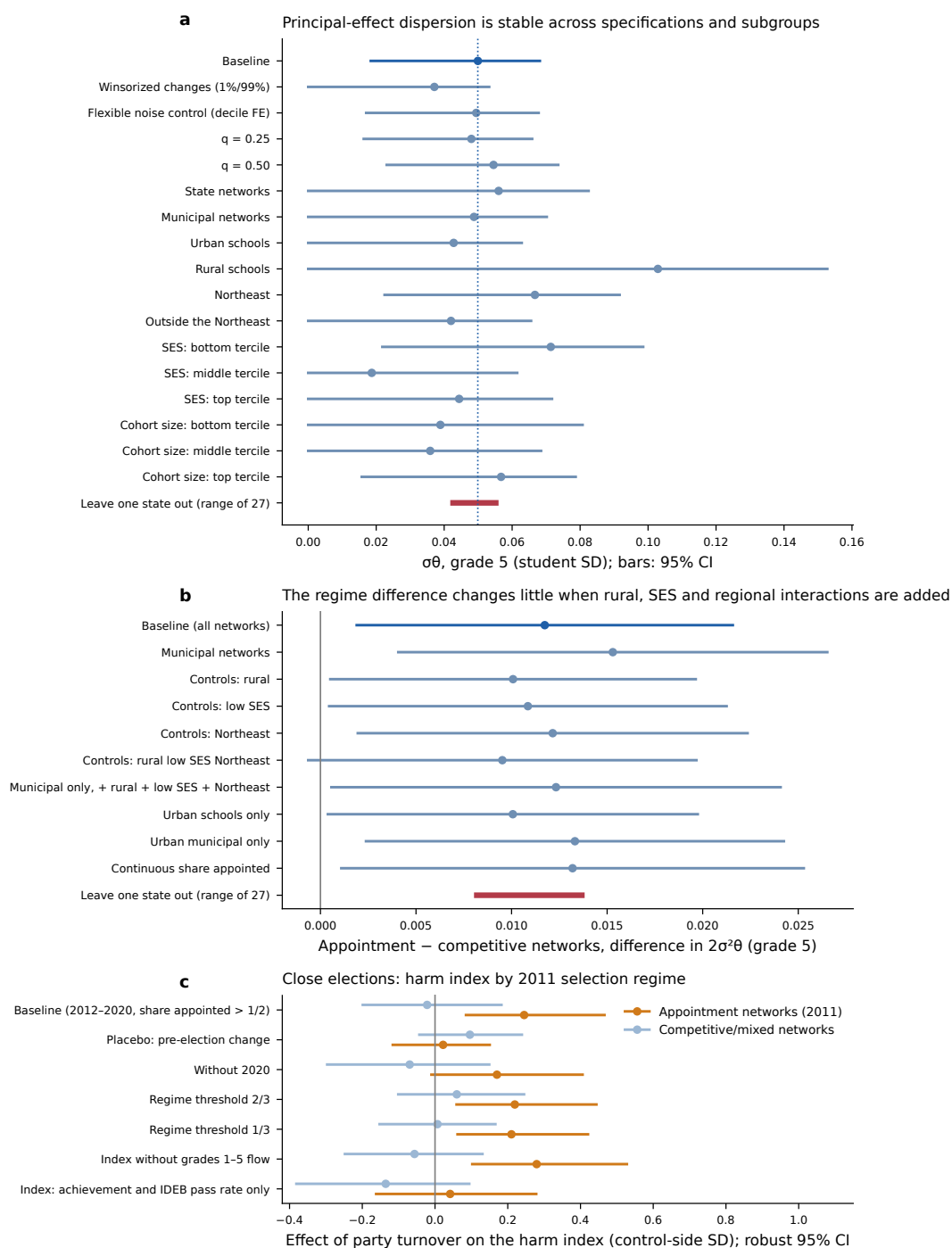
A patronage model of the kind formalized in Section 3.2, and documented for Brazilian municipal bureaucracies by Colonnelli, Prem & Teso (2020), accounts for both patterns. When appointment weighs attributes unrelated to management, the distribution of management effects among appointed principals is wide, and successions, which occur disproportionately in declining schools, produce more dispersed and on average worse outcomes. Where selection is competitive, principals appear approximately interchangeable with respect to achievement, which is also what Bartanen, Husain & Liebowitz (2024) conclude for U.S. systems. Whether principals matter may thus depend on the selection technology of the system in which they work. Selection regimes are not randomly assigned, however, and we interpret the difference as a robust pattern of heterogeneity rather than as the causal effect of changing a network’s rule. The next subsection examines whether the pattern could reflect other characteristics of appointment networks.

6.6 Robustness and heterogeneity

Table 6, Table 7 and Figure 4 collect the robustness evidence for the two primary second-moment results, the grade-5 standard deviation and its difference between regimes.

The estimated standard deviation is insensitive to the treatment of sampling noise and to the influence of individual states. Replacing the linear control for expected sampling variance with decile fixed effects leaves the grade-5 estimate at 0.049 (Table 6). Winsorizing two-year changes at the 1st and 99th percentiles lowers it to 0.037 (upper bound 0.053) and raises the grade-9 estimate to 0.024 (upper bound 0.044). That the grade-5 estimate falls when the tails are trimmed indicates that part of the excess variance comes from a minority of successions with large effects, which is what a distribution of principal effects with some very weak and some very strong managers would generate. Dropping each state in turn leaves the grade-5 estimate between 0.043, when Maranhão

Figure 4—Robustness of the main results



Notes: (a) σ_{θ} in grade 5 across specifications and subgroups (Table 6), with the range of 27 leave-one-state-out estimates; dotted line at the baseline. (b) Appointment-minus-competitive difference in $2\sigma_{\theta}^2$ (Table 7, Panel B). (c) Effect of party turnover in close elections on the summary harm index, by the 2011 selection regime (Table 10). Bars: 95% confidence intervals (robust bias-corrected in panel c).

is excluded, and 0.055, when Santa Catarina is excluded, with upper bounds between 0.063 and 0.073, while the grade-9 estimate stays between 0 and 0.012. Because the asymptotic distribution of a clustered variance estimator may be a poor guide in finite samples, we also report randomization inference that permutes succession groups across schools within state $\times$ dependency $\times$ location cells. Across 300 permutations, the one-sided p-value is 0.017 in grade 5, where the observed $2\sigma_\theta^2$ of 0.0050 compares with a null mean of -0.0002 and a null standard deviation of 0.0023, and 0.64 in grade 9.

Subgroup estimates reveal where the dispersion is larger, without overturning the national picture. Estimated separately, grade-5 standard deviations range from 0.019 in the middle tercile of socioeconomic status to 0.103 in rural schools, where the estimate is imprecise (upper bound 0.153, based on 4,941 schools). Point estimates are also larger in the bottom tercile of socioeconomic status (0.071) and in the Northeast (0.067). Pooled interaction tests, however, reject equality in none of these cases at the 5% level (Table 7, Panel A); the closest are the comparisons by socioeconomic status ($p = 0.086$) and rurality ($p = 0.167$). The estimate is not an artifact of small cohorts, whose school means are noisiest: it is 0.057 (upper bound 0.079) in the top tercile of cohort size and 0.039 in the bottom tercile. In grade 9, no subgroup estimate or interaction is significant.

These subgroup patterns raise an obvious concern about the regime difference. Appointment is more common in rural, poorer and northeastern networks, where the subgroup estimates are also larger, so the regime could simply proxy for disadvantage. Table 7, Panel B addresses this concern by adding interactions of the succession probability with each of these characteristics. The regime difference is essentially unchanged when each competing interaction is added (between 0.0101 and 0.0122; $p \leq 0.041$). When all three enter together, it is 0.0095 ($p = 0.066$), and within municipal networks with all three it is 0.0123 ($p = 0.040$). None of the competing interactions is itself significant ($p \geq 0.18$). The difference also holds among urban schools (0.0101; $p = 0.042$) and urban municipal schools (0.0133; $p = 0.017$), and when the classification is replaced by the continuous leave-one-out share of appointed principals (0.0132; $p = 0.033$). Across the 27 samples that drop one state at a time, it ranges from 0.0082 to 0.0137 and is significant at the 5% level in 25; the largest p-value, 0.104, again arises when Maranhão is excluded.

Randomization inference is less favorable to the regime difference than conventional inference. Permuting regime labels across networks within state $\times$ dependency cells yields a one-sided p-value of 0.053, because the permutation distribution is centered slightly above zero (at 0.0027). We read this evidence as follows. Dispersion tracks the selection rule rather than rurality, poverty or region, and the difference is stable across specifications and samples, but its statistical strength is moderate and the rule itself is not randomly assigned.

Table 6—Robustness of the standard deviation of principal-by-school effects σ_θ by specification and subgroup (student SD; $q = 0.33$ unless stated)

Specification / subgroup	Grade 5: $2\sigma_\theta^2$ (s.e.)	σ_θ	95% upper bound	Schools	Grade 9: $2\sigma_\theta^2$ (s.e.)	σ_θ	95% upper bound	Schools
Baseline	0.0050 (0.0022)	0.050	0.068	37,264	−0.0008 (0.0021)	0.000	0.041	28,567
Winsorized changes (1%/99%)	0.0028 (0.0015)	0.037	0.053	37,264	0.0012 (0.0014)	0.024	0.044	28,567
Flexible noise control (decile FE)	0.0049 (0.0022)	0.049	0.068	37,264	−0.0008 (0.0021)	0.000	0.041	28,567
$q = 0.25$	0.0046 (0.0021)	0.048	0.066	37,264	−0.0008 (0.0020)	0.000	0.040	28,567
$q = 0.50$	0.0059 (0.0025)	0.055	0.074	37,264	−0.0009 (0.0024)	0.000	0.044	28,567
State networks	0.0063 (0.0037)	0.056	0.083	10,452	−0.0013 (0.0027)	0.000	0.045	16,347
Municipal networks	0.0048 (0.0026)	0.049	0.070	26,812	0.0002 (0.0035)	0.009	0.060	12,220
Urban schools	0.0037 (0.0022)	0.043	0.063	32,323	−0.0012 (0.0022)	0.000	0.039	24,984
Rural schools	0.0212 (0.0130)	0.103	0.153	4,941	0.0246 (0.0218)	0.111	0.183	3,583
Northeast	0.0089 (0.0040)	0.067	0.092	11,992	−0.0057 (0.0038)	0.000	0.029	8,714
Outside the Northeast	0.0035 (0.0026)	0.042	0.066	25,272	0.0013 (0.0025)	0.025	0.056	19,853
SES: bottom tercile	0.0102 (0.0047)	0.071	0.099	12,364	−0.0023 (0.0038)	0.000	0.051	9,474
SES: middle tercile	0.0007 (0.0035)	0.019	0.061	12,406	0.0008 (0.0036)	0.020	0.063	9,488
SES: top tercile	0.0039 (0.0032)	0.044	0.072	12,474	0.0005 (0.0036)	0.015	0.061	9,521
Cohort size: bottom tercile	0.0030 (0.0051)	0.039	0.081	12,276	−0.0009 (0.0056)	0.000	0.071	9,485
Cohort size: middle tercile	0.0026 (0.0035)	0.036	0.069	12,480	−0.0037 (0.0030)	0.000	0.033	9,531
Cohort size: top tercile	0.0064 (0.0030)	0.057	0.079	12,508	0.0022 (0.0029)	0.033	0.063	9,551

Notes: All specifications use the variance-of-changes estimator with groups T, C, D and S, cell $\times$ window and group $\times$ window effects, and errors clustered by municipality. “Winsorized” trims two-year changes at the 1st and 99th percentiles of each cell; “Flexible noise control” replaces the linear control for expected sampling variance by decile fixed effects; SES and cohort-size terciles are defined on school means in 2007.

Table 7—Heterogeneity tests and the regime difference net of competing explanations*Panel A. Pooled interaction tests: $\pi \times \text{subgroup}$ (difference in $2\sigma_\theta^2$)*

Contrast	Grade 5: σ_θ reference / subgroup	Difference (s.e.)	p	Grade 9: σ_θ reference / subgroup	Difference (s.e.)	p
Rural vs urban	0.043 / 0.104	0.0180 (0.0130)	0.167	0.000 / 0.126	0.0330 (0.0230)	0.152
Municipal vs state networks	0.056 / 0.049	−0.0015 (0.0045)	0.740	0.000 / 0.009	0.0014 (0.0046)	0.753
Northeast vs other regions	0.042 / 0.066	0.0053 (0.0048)	0.269	0.025 / 0.000	−0.0070 (0.0046)	0.122
SES below vs above median	0.027 / 0.067	0.0074 (0.0043)	0.086	0.023 / 0.000	−0.0036 (0.0040)	0.373
Cohort size below vs above median	0.054 / 0.043	−0.0022 (0.0045)	0.625	0.022 / 0.000	−0.0033 (0.0043)	0.443

Panel B. Appointment-minus-competitive difference in $2\sigma_\theta^2$, grade 5, with competing interactions

Specification	Difference (s.e.)	p	Competing $\pi \times Z$ coefficients	Schools
Regime only (baseline, Table 5)	0.0117 (0.0050)	0.020	—	28,053
+ $\pi \times \text{rural}$	0.0101 (0.0049)	0.039	rural 0.0159 ($p = 0.30$)	28,053
+ $\pi \times \text{low SES}$	0.0109 (0.0053)	0.041	low SES 0.0030 ($p = 0.56$)	28,038
+ $\pi \times \text{Northeast}$	0.0122 (0.0052)	0.020	Northeast −0.0007 ($p = 0.90$)	28,053
+ all three	0.0095 (0.0052)	0.066	rural 0.0221 ($p = 0.18$); low SES 0.0030 ($p = 0.67$); Northeast −0.0017 ($p = 0.82$)	28,038
Municipal networks only, + all three	0.0123 (0.0060)	0.040	rural 0.0247 ($p = 0.20$); low SES 0.0043 ($p = 0.61$); Northeast −0.0020 ($p = 0.82$)	20,482
Urban schools only	0.0101 (0.0049)	0.042	—	24,023
Urban municipal schools only	0.0133 (0.0056)	0.017	—	16,780
Continuous leave-one-out share appointed ($\pi \times \text{share}$)	0.0132 (0.0062)	0.033	—	—
Leave one state out (27 estimates)	0.0082 to 0.0137	max 0.104	$p < 0.05$ in 25 of 27	—
Randomization inference (300 permutations of regime within state $\times$ dependency)	0.0117	0.053 (one-sided)	null mean 0.0027, s.d. 0.0053	—

Notes: All specifications use the variance-of-changes estimator with groups T, C, D and S, cell $\times$ window and group $\times$ window effects, and errors clustered by municipality. Panel A: a single regression with π and $\pi \times Z$; the reference and subgroup σ_θ are implied by the two coefficients. Panel B: $Z = \text{rural, below-median SES, Northeast}$; the continuous specification replaces the regime indicator by the network’s leave-one-out share of appointed principals (all networks). Leave-one-state-out: the largest p (0.104) arises when Maranhão is excluded.

6.7 Leadership and management indices

A large literature relates survey measures of management practices and leadership to school outcomes, often in cross-section. Our design allows a simple test of whether such indices measure the causal contribution of the current principal. If teacher-rated leadership measured in 2011 captured that contribution, three predictions should hold. The index should not predict changes in achievement in stable schools, where the principal did not change during the measurement window; it should predict post-arrival changes in succession schools; and it should not predict changes that preceded the successor’s arrival. The data reject all three (Appendix Table A3). In stable schools, one standard deviation of leadership is associated with 0.030 SD (s.e. 0.005) of growth between 2005 and 2009, a placebo window in which the rated principal was already in post and nothing changed, and with only 0.009 (s.e. 0.005) between 2009 and 2011. Its interaction with succession is −0.013 (s.e. 0.006) in the placebo window and −0.005 (0.006), −0.003 (0.007) and +0.006

(0.007) for changes to 2011, 2013 and 2015. The principal-reported practice index follows the same pattern, with a placebo coefficient of 0.014 and interactions between -0.006 and $+0.007$, and the grade-9 results are similar.

The simplest interpretation is reverse causality or a halo effect: teachers rate principals of improving schools more favorably, and principals of successful schools report more structured practices. This does not imply that management practices are irrelevant to learning, and it is consistent with evidence that people management is associated with school value added when productivity is measured more carefully (Lemos, Muralidharan, and Scur 2024). It does imply that cross-sectional correlations between management indices and achievement, even with municipality fixed effects, do not identify the effect of management in this setting.

6.8 Inputs and school flow

If new principals do not change measured learning on average, they might still change inputs or other outcomes. In a balanced panel of 27,705 schools observed from 2007 to 2015, successions do not change enrollment (log enrollment coefficients between -0.005 and $+0.005$), teachers per class (between -0.010 and $+0.011$) or library availability. Class size rises by 0.10–0.16 students from 2014 onward, and the presence of a computer laboratory falls by 1.0–2.6 percentage points. The decline in laboratories, however, begins in 2010 also among schools whose new principal arrived only between November 2010 and November 2011, that is, before that principal took office, given that the census reference date is in May. This timing points to selection rather than to an effect of the principal.

School flow in the early grades shows a more encouraging pattern. Comparing 12,340 group-T schools offering grades 1–5 with 7,779 stable schools year by year, pass rates are flat before the succession (2007: -0.16 , s.e. 0.15; 2008: $+0.04$, s.e. 0.14) and rise gradually afterward, reaching $+0.44$ percentage points in 2013 (s.e. 0.13) on a base of 88.1%. Dropout falls by about 0.15 points from 2012 on a base of 2.2%. Pass rates and age-grade distortion in grades 6–9 show pre-trends and are not interpreted. New principals thus appear to reduce grade retention modestly without changing measured learning, a pattern we regard as descriptive given the selection documented in Section 6.1.

7 Political Turnover, Principals and Teachers

7.1 Party turnover replaces municipal principals

The 2008 cycle is the only one for which the timing of principal arrivals can be observed, because the 2011 questionnaire records the tenure of the principal in post three years after the election. In the triple-difference specification with municipality fixed effects (Table 8, Panel A), a change in the mayor's party raises the probability that the municipal principal in post in 2011 arrived after November 2006 by 9.1 percentage points (s.e. 1.5; $F \approx 35$), on a base of 62.7% in municipalities without turnover. The effect is concentrated in the arrival window that contains the January 2009 inauguration: +6.0 points (s.e. 1.7) for arrivals between November 2006 and November 2009, against +3.1 points for arrivals between 2009 and 2011. Party turnover also raises the share of principals who report having been appointed by politicians, by 4.0 points (s.e. 1.8).

Prediction P3 holds: turnover is confined to networks where the post is discretionary. To test it we classify municipalities by a predetermined regime, the share of appointees among municipal principals who were already in post before November 2006. Party turnover raises arrivals in 2006–09 by 7.8 points (s.e. 2.3) in appointment networks and by -2.4 points (s.e. 2.4) elsewhere, a difference of 10.4 points (s.e. 3.4; $p = 0.002$). Because the regime measure is built from surviving principals, turnover itself shifts it: municipalities with turnover show 4.8 points less predetermined appointment (s.e. 1.9), since some earlier appointees had been replaced. This misclassification moves appointment networks into the comparison group and should therefore attenuate the difference between regimes rather than create it.

7.2 Triple differences and retrospective voting

In the 2008 cycle, the triple difference suggests that turnover harms learning. Relative to state schools in the same municipality, grade-5 achievement in municipal schools falls by 0.047 SD in 2009 (s.e. 0.018) after party turnover, with a flat pre-period (2005: -0.010 , s.e. 0.020), and age-grade distortion in grades 6–9 rises by 0.83 percentage points on average over 2009–12 (s.e. 0.38) on a base of 41%. Two features of these estimates counsel caution. The implied Wald ratio, the reduced form divided by the first stage, is about -0.37 SD per induced change of principal, which is implausibly large given a standard deviation of principal effects near 0.05 and an average succession effect near zero. And the effect does not recur in later cycles.

Longer pre-periods in the 2012, 2016 and 2020 cycles reveal why (Figure 5). In the 2016 cycle, the grade 6–9 pass rate in municipal schools of municipalities about to experience party turnover was already falling relative to state schools: the estimates are +0.93, +1.26 and +0.94 percentage points in years -4 to -2 , against zero in the year before the election. Grade-9 achievement declines

Table 8—Party turnover, principal replacement and school outcomes: triple differences

Panel A. First stage, 2008 cycle (principals in 2011; municipality, state $\times$ network and network $\times$ location fixed effects)

Outcome	All municipalities: β (s.e.)	Incumbent margin < 10 pp: β (s.e.)	Mean without turnover
Arrived after Nov 2006	+0.091 (0.015)	+0.143 (0.055)	0.627
Arrived Nov 2006 – Nov 2009	+0.060 (0.017)	+0.178 (0.054)	0.317
Arrived Nov 2009 – Nov 2011	+0.031 (0.014)	−0.035 (0.055)	0.309
Appointed by politicians	+0.040 (0.018)	+0.103 (0.055)	0.248
Arrived Nov 2006 – Nov 2009, appointment minus other networks	+0.104 (0.034)		

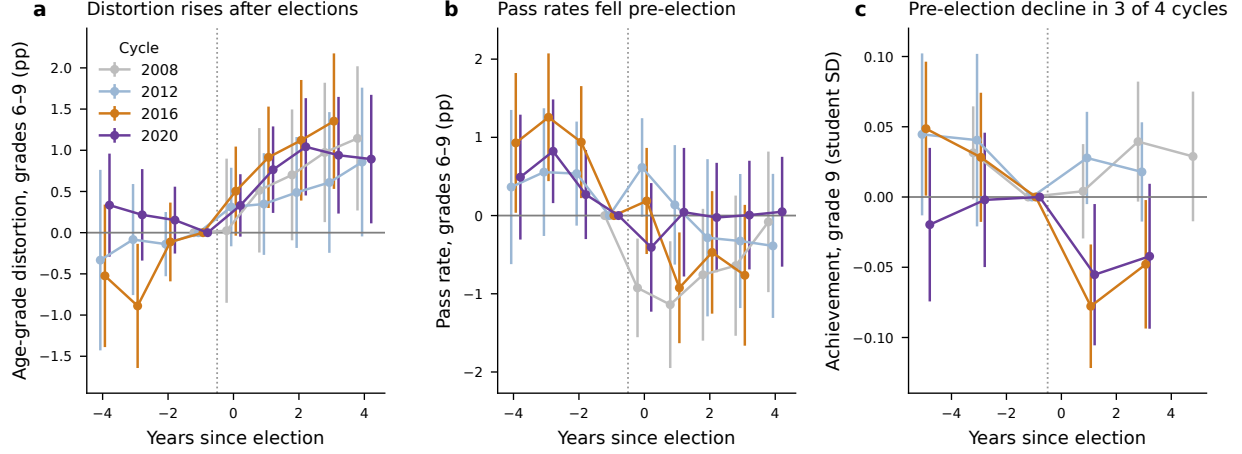
Panel B. Stacked cycles, post-election average (years +1 to +3) relative to the year before the election

Outcome	Cycles	Unadjusted	Adjusted for linear pre-trend	Pre-trend of treated (per year or wave)
Age-grade distortion, grades 6–9 (pp)	2012, 2016, 2020	+0.82 (0.19)	+0.58 (0.34)	+0.08 (0.09)
Age-grade distortion, grades 6–9 (pp)	2012, 2016	+0.73 (0.24)	+0.21 (0.47)	+0.18 (0.13)
Pass rate, grades 6–9 (pp)	2012, 2016, 2020	−0.26 (0.20)	+0.21 (0.38)	−0.22 (0.09)
Repetition, grades 6–9 (pp)	2012, 2016, 2020	+0.19 (0.20)	−0.42 (0.37)	+0.25 (0.08)
Achievement, grade 5 (student SD)	2008–2020	−0.018 (0.009)	−0.005 (0.016)	−0.003 (0.003)
Achievement, grade 9 (student SD)	2008–2020	−0.004 (0.009)	+0.024 (0.019)	−0.010 (0.005)

Notes: Panel A: $n = 49,459$ schools in 4,196 municipalities (all) and 9,020 schools in 1,005 municipalities (margin below 10 points, with separate margin slopes on each side). Panel B: school, municipality $\times$ year, state $\times$ network $\times$ year and network $\times$ location $\times$ year fixed effects specific to each cycle; errors clustered by municipality. Flow window −4 to +3 years; achievement waves −5, −3, −1, +1, +3.

before the election in three of the four cycles. When cycles are stacked with cycle-specific fixed effects (Table 8, Panel B), party turnover is preceded by a significant relative decline of the municipal network: the treated group’s pre-election trend is -0.22 points per year in the grade 6–9 pass rate (s.e. 0.09), $+0.25$ in repetition (s.e. 0.08) and -0.010 SD per test wave in grade-9 achievement (s.e. 0.005).

Figure 5—Pre-election trends by electoral cycle



Notes: Dynamic triple difference (party turnover $\times$ municipal school), reference year before the election (dotted line between -1 and 0). (a) Age-grade distortion and (b) pass rates in grades 6–9 (percentage points); the 2008 cycle has a single pre-election year in the flow panel. (c) Grade-9 achievement (student SD; biennial tests). 95% confidence intervals; errors clustered by municipality.

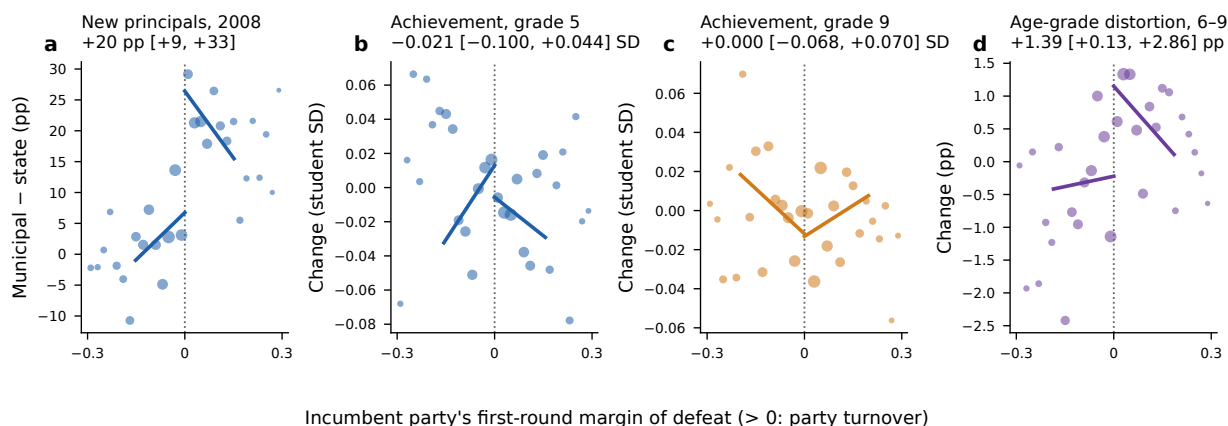
Once post-election effects are measured as deviations from this linear pre-trend, none of them is robust. The increase in age-grade distortion falls from 0.82 (s.e. 0.19) to 0.58 (s.e. 0.34), and to 0.21 (s.e. 0.47) when the pandemic-affected 2020 cycle is excluded, and the grade-5 achievement effect falls from -0.018 (s.e. 0.009) to -0.005 (s.e. 0.016). Voters appear to remove incumbents where municipal schools are deteriorating, in line with the evidence on retrospective voting in Brazilian municipalities (Ferraz and Finan 2008). The resulting correlation between turnover and pre-existing trends violates the parallel-trends assumption and would lead a researcher relying on the triple difference to attribute to the change of government a decline that had begun before it.

7.3 Close elections

The stacked regression discontinuity addresses this endogeneity (Figure 6; Table 9). We first assess its validity. Continuity of the density of the running variable cannot be rejected when cycles are pooled ($T = 0.05$; $p = 0.96$). By cycle, the test rejects at the 5% level in 2008 ($p = 0.029$) and 2012 ($p = 0.010$), but with opposite signs, which is not suggestive of systematic sorting of incumbents on one side of the cutoff. Of 24 balance tests on pre-election levels and pre-election changes of twelve

outcomes, one rejects at 5% (the pre-election change in grade 1–5 age-grade distortion, $p = 0.030$) and one is borderline (the pre-election level of grade 1–5 repetition, $p = 0.051$), as expected under the null. In particular, pre-election changes in achievement and pass rates are balanced at the cutoff: the deterioration that invalidates the triple difference is absent in close elections.

Figure 6—Stacked regression discontinuity in close mayoral elections



Notes: Means in 2-point bins of the incumbent party's first-round margin of defeat (dot size proportional to the number of municipalities) and local linear fits with a triangular kernel within the MSE-optimal bandwidth on each side. (a) 2008 first stage: municipal-minus-state difference in the share of principals who arrived after November 2006 (SAEB 2011). (b–d) Post-election change in the municipal-minus-state contrast, four cycles pooled, net of cycle means. Panel titles report the estimate and the robust 95% confidence interval.

The first stage is strong. In close 2008 elections, party turnover raises the municipal-minus-state difference in the share of principals who arrived after November 2006 by 20.3 percentage points (robust 95% CI: 8.9 to 33.2; $p = 0.001$). As in the triple difference, the increase is concentrated in the window containing the inauguration, with +17.3 points (3.7 to 30.4) for arrivals in 2006–09 against +2.4 points (–11.3 to 17.4) for arrivals in 2009–11. The effect on the share of politically appointed principals is +5.6 points (–9.2 to 18.1), smaller and imprecise, which is consistent with new mayors replacing appointees with other appointees rather than changing the mode of selection.

Consistent with P4, party turnover in close elections has no detectable effect on average achievement. The estimate is –0.021 SD in grade 5 (CI –0.100 to +0.044) and 0.000 in grade 9 (CI –0.068 to +0.070), so declines larger than 0.10 student SD in grade 5 and effects larger than 0.07 SD in either direction in grade 9 can be ruled out. For the 2008 cycle alone, the grade-5 point estimate (–0.047; CI –0.168 to +0.072) coincides with the triple difference but cannot be distinguished from zero, and the estimates for the other cycles range from –0.085 to +0.015. Of the twelve outcomes in Table 9, two flow coefficients are individually significant, and neither survives an adjustment for multiple testing (Benjamini–Hochberg $q \geq 0.17$) or the robustness checks in Appendix Table A4. The increase of 1.39 points in grade 6–9 age-grade distortion is driven by the 2020 cycle, in which the estimate alone is +4.67; without 2020 it is +0.88 ($p = 0.25$), and net of its

pre-election change +1.44 ($p = 0.20$). The decline of 1.32 points in grade 1–5 repetition is stable across bandwidths, but the pre-election level of repetition is imbalanced in the opposite direction (+0.91; $p = 0.051$), and net of the pre-election change the effect is -0.54 ($p = 0.49$), as mean reversion would imply. Estimating the discontinuity at false cutoffs of ± 5 and ± 10 points, using observations on one side of the true cutoff only, produces one rejection at the 5% level in 20 tests for the headline outcomes (Appendix Table A6).

Table 9—Stacked regression discontinuity in close elections, 2008–2020

Outcome (municipal – state change)	Estimate [robust 95% CI]	Robust p	BH q	Municipality-cycles, left / right
Achievement, grade 5 (student SD)	–0.021 [–0.100, +0.044]	0.447	0.802	974 / 1,004
Achievement, grade 9 (student SD)	+0.000 [–0.068, +0.070]	0.975	0.975	1,102 / 1,122
Pass-rate indicator, grade 5 (0–1)	+0.005 [–0.010, +0.023]	0.456	0.802	1,228 / 1,211
Pass-rate indicator, grade 9 (0–1)	–0.004 [–0.026, +0.018]	0.746	0.814	1,354 / 1,371
Age-grade distortion, grades 6–9 (pp)	+1.39 [+0.13, +2.86]	0.032	0.191	1,984 / 1,959
Pass rate, grades 6–9 (pp)	–0.28 [–1.86, +1.19]	0.668	0.802	1,853 / 1,823
Repetition, grades 6–9 (pp)	+0.33 [–0.94, +1.57]	0.621	0.802	2,030 / 2,008
Dropout, grades 6–9 (pp)	+0.13 [–0.63, +1.10]	0.596	0.802	1,683 / 1,671
Age-grade distortion, grades 1–5 (pp)	–0.21 [–1.80, +1.15]	0.664	0.802	1,575 / 1,510
Pass rate, grades 1–5 (pp)	+0.90 [–0.26, +2.28]	0.118	0.464	1,377 / 1,337
Repetition, grades 1–5 (pp)	–1.32 [–2.64, –0.29]	0.014	0.172	1,209 / 1,176
Dropout, grades 1–5 (pp)	+0.37 [–0.16, +0.98]	0.155	0.464	1,743 / 1,667
<i>First stage (2008): arrived after Nov 2006 (pp)</i>	<i>+20.3 [+8.9, +33.2]</i>	<i>0.001</i>		<i>772 / 620</i>

Notes: Unit is municipality $\times$ cycle; sample restricted to municipalities where the incumbent party ran and no runoff took place (12,527 municipality-cycles). Local linear, triangular kernel, MSE-optimal bandwidth, bias-corrected robust inference, cycle indicators as covariates, errors clustered by municipality. q: Benjamini–Hochberg adjustment across the twelve outcomes.

7.4 Heterogeneity by selection regime

The pooled estimate averages over networks where the mayor controls the principalship and networks where the mayor does not. Prediction P5 implies that any effect of turnover should appear in the former, and the first stage confirms that replacement concentrates there. Testing this implication outcome by outcome would multiply the number of comparisons, so we summarize the ten outcomes in a single index (Kling, Liebman, and Katz 2007). The index averages the standardized post-election changes in achievement and school flow, each signed so that higher values mean worse outcomes:

$$H_{mc} = \frac{1}{|\mathcal{K}_{mc}|} \sum_{k \in \mathcal{K}_{mc}} s_k \frac{\Delta Y_{mc}^k - \mu_0^k}{\sigma_0^k}, \quad (8)$$

where ΔY_{mc}^k is the contrast (6) for outcome k , $s_k \in \{-1, +1\}$ orients each outcome, μ_0^k and σ_0^k are its mean and standard deviation on the control side of the cutoff, and $\mathcal{K}_{mc}$ is the set of available outcomes, of which we require at least three. We split municipalities by the share of municipal principals who reported being appointed in SAEB 2011, a measure that is predetermined for the 2012, 2016 and 2020 elections (Table 10; Figure 4c).

Pooled across networks, party turnover does not move the index (+0.012; CI -0.066 to $+0.097$), and the index is balanced before the election. In appointment networks, by contrast, turnover raises the index by 0.245 standard deviations (robust CI 0.085 to 0.466; $p = 0.005$), with a balanced pre-election change (+0.022; $p = 0.80$). In competitive or mixed networks the effect is -0.022 ($p = 0.94$). The estimate for appointment networks is positive in every variant we examined and significant at the 5% level in four of six: it is 0.21–0.22 with regime thresholds of one or two thirds and 0.28 when the grade 1–5 flow components are dropped, and it falls to 0.17 ($p = 0.062$) when the 2020 cycle is excluded. The difference between regimes is less precise, at 0.267 (s.e. 0.137; $p = 0.052$) in the baseline, between 0.160 and 0.336 across variants, and significant at the 5% level in one of six. No placebo cutoff produces a significant effect on the index in either regime (Appendix Table A6).

The harm operates through grade progression rather than test scores. An index restricted to achievement and the pass-rate indicators gives +0.042 ($p = 0.60$). Among the components in appointment networks (Appendix Table A7), grade 6–9 age-grade distortion rises by 2.98 points (CI 0.84 to 5.99), the grade 6–9 pass rate falls by 2.10 points ($p = 0.125$), and repetition and dropout rise by 1.44 and 0.80 points, neither significantly. The achievement estimates in grades 5 and 9 are negative (-0.049 and -0.043 SD) but imprecise. In competitive or mixed networks, the grade 6–9 flow estimates have mixed signs, and none is significant.

7.5 Teachers: the scope of the personnel shock

The flow effects in appointment networks could reflect the replacement of principals or a broader reshuffling of school staff. Prediction P5 does not distinguish between the two, and the magnitudes discussed in Section 3.3 make the first explanation unlikely on its own. Public data do not allow principal turnover to be measured by municipality after 2011, but they do measure the teaching staff (Section 4.4). We therefore estimate the same stacked regression discontinuity for INEP’s teacher regularity indicator, which is available for the 2016 and 2020 cycles, and for the contract composition of teachers in grades 1–9, which is available for all three cycles in the regime split (Table 11 and Appendix Table A8; Figure 7).

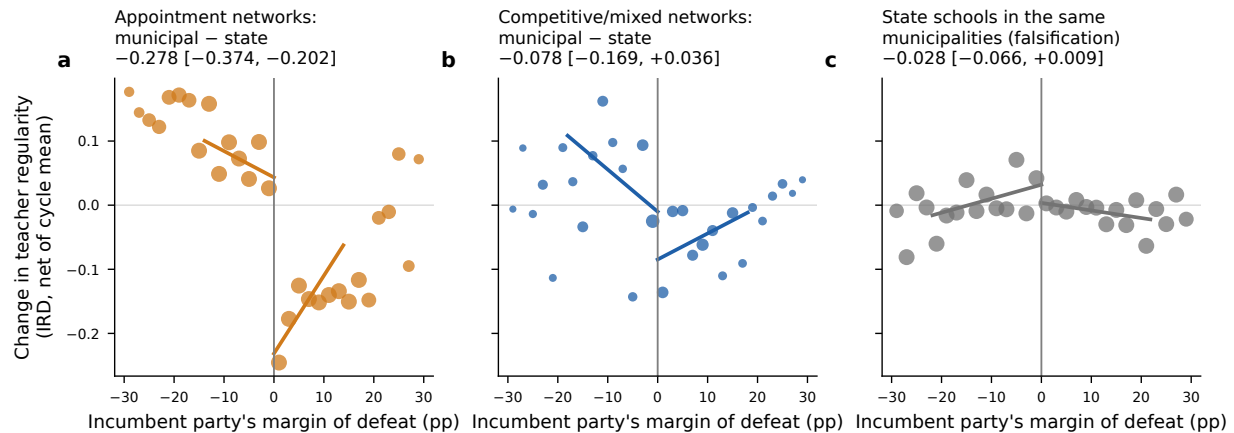
Party turnover in close elections sharply lowers teacher regularity in the municipal network. The municipal-minus-state indicator falls by 0.199 points (robust CI -0.259 to -0.142), about

Table 10—Close elections by selection regime: summary harm index

Specification	Appointment networks: estimate [robust 95% CI]		p	Left / right		Competitive or mixed: estimate [robust 95% CI]		p	Left / right		Difference (s.e.)		p
Baseline: 2012–2020 cycles, share appointed > 1/2													
Placebo: pre-election change of the index	+0.245 [+0.085, +0.466]	0.005	499 / 573	–0.022 [–0.199, +0.183]	0.935	227 / 288	+0.267 (0.137)	0.052					
Without the 2020 cycle	+0.022 [–0.116, +0.150]	0.801	673 / 789	+0.096 [–0.043, +0.239]	0.173	253 / 310	–0.074 (0.099)	0.456					
Regime threshold: share appointed > 2/3	+0.170 [–0.010, +0.406]	0.062	386 / 468	–0.070 [–0.296, +0.149]	0.516	159 / 242	+0.240 (0.155)	0.123					
Regime threshold: share appointed > 1/3	+0.219 [+0.059, +0.444]	0.010	479 / 545	+0.060 [–0.101, +0.245]	0.415	329 / 416	+0.160 (0.132)	0.227					
Index without grade 1–5 flow components	+0.210 [+0.062, +0.421]	0.008	560 / 645	+0.007 [–0.152, +0.166]	0.933	177 / 223	+0.203 (0.122)	0.096					
Index of achievement and pass-rate indicators only	+0.280 [+0.102, +0.528]	0.004	440 / 498	–0.056 [–0.248, +0.130]	0.543	184 / 233	+0.336 (0.145)	0.021					
<i>All networks, four cycles (2008–2020)</i>	+0.042 [–0.162, +0.278]	0.604	290 / 352	–0.136 [–0.381, +0.094]	0.236	130 / 173	+0.177 (0.165)	0.284					
<i>All networks, four cycles, placebo</i>	+0.012 [–0.066, +0.097]	0.714	2,413 / 2,336										
	+0.030 [–0.054, +0.127]	0.424	1,402 / 1,519										

Notes: The harm index is the mean of up to ten standardized post-election changes in the municipal-minus-state contrast — grade 5 and grade 9 achievement, and age-grade distortion, pass, repetition and dropout rates in grades 1–5 and 6–9 — each signed so that higher values mean worse outcomes and standardized by the mean and standard deviation on the control side of the cutoff (incumbent party re-elected); at least three components are required. Regime: share of municipal principals in SAEB 2011 who report being appointed (technical, political or other appointment), for 3,634 municipalities with at least two respondents; 74.5% exceed one half. Because the regime is measured in 2011, the 2008 cycle is excluded from the regime split. Specification as in Table 9. Difference: independent-samples test with robust standard errors.

Figure 7—Party turnover and teacher regularity in close elections, 2016 and 2020



Notes: Change in the teacher regularity indicator between May of the election year and the mean of the two following censuses, net of cycle means, in 2-point bins of the incumbent party's margin of defeat (dot size proportional to the number of municipalities), with local linear fits within the MSE-optimal bandwidth. (a–b) Municipal-minus-state contrast by the 2011 selection regime. (c) State schools in the same municipalities, whose staff do not answer to the mayor. Panel titles report the estimate and the robust 95% confidence interval.

0.55 standard deviations of its change on the control side. The effect is concentrated in appointment networks, where it is -0.278 (CI -0.374 to -0.202), whereas it is smaller and not significant in competitive or mixed networks (-0.078 ; $p = 0.20$). The difference between regimes is significant (-0.199 , s.e. 0.068 ; $p = 0.004$), and remains so with regime thresholds of one and two thirds ($p = 0.012$ and 0.017). The drop appears in the first census after the inauguration and holds separately in 2016 and 2020, and it is stable across bandwidths from half to twice the optimal one, while no false cutoff produces a significant estimate. The falsification test is informative. State schools in the same municipalities, whose teachers do not answer to the mayor, show no change (-0.028 ; $p = 0.14$), so the effect is confined to the network that the mayor controls.

One diagnostic requires care. In the pooled sample, the indicator rose before the 2020 election in municipalities where the incumbent party went on to lose narrowly ($+0.125$; $p = 0.004$), which raises the possibility that the post-election decline reflects mean reversion. The pre-election change is not significant within either regime, however, and netting it out makes the 2020 estimate larger rather than smaller (-0.304 overall and -0.299 in appointment networks). The pre-election level of the indicator is balanced.

What changes is the identity of the teachers, not their number or their contracts (Appendix Table A8). The municipal-minus-state share of teachers on temporary contracts moves by $+0.010$ (CI -0.021 to $+0.034$), with no difference between regimes, and the total number of municipal teachers is unchanged ($+0.7\%$; $p = 0.57$). The number of permanent municipal teachers falls slightly (-3.1% ; $p = 0.014$), without a regime pattern. The combination of a lower regularity indicator with stable headcounts and contract shares is what one would expect if temporary teachers

Table 11—Party turnover in close elections and teacher regularity

Teacher regularity indicator (IRD, school mean, 0–5 scale); municipal — state contrast unless stated

Specification	All	p	Appointment	p	Competitive or mixed	p	Difference (s.e.)	p
Baseline: mean(c+1, c+2) – c, 2016 and 2020 cycles	–0.199 [–0.259, –0.142]	< 0.001	–0.278 [–0.374, –0.202]	< 0.001	–0.078 [–0.169, +0.036]	0.204	–0.199 (0.068)	0.004
First post-election year only: (c+1) – c	–0.191 [–0.237, –0.141]	< 0.001	–0.250 [–0.337, –0.182]	< 0.001	–0.054 [–0.130, +0.048]	0.365	–0.196 (0.060)	0.001
2016 cycle only	–0.212 [–0.298, –0.143]	< 0.001	–0.291 [–0.415, –0.191]	< 0.001	–0.078 [–0.198, +0.068]	0.338	–0.213 (0.089)	0.016
2020 cycle only	–0.188 [–0.267, –0.091]	< 0.001	–0.257 [–0.388, –0.142]	< 0.001	–0.064 [–0.215, +0.124]	0.599	–0.193 (0.107)	0.071
2020 cycle, net of pre-election change (2020 – 2018)	–0.304 [–0.436, –0.178]	< 0.001	–0.299 [–0.480, –0.111]	0.002	–0.148 [–0.445, +0.188]	0.425	–0.151 (0.187)	0.419
Placebo: pre-election change, 2020 – 2018	+0.125 [+0.044, +0.234]	0.004	+0.042 [–0.068, +0.178]	0.382	+0.081 [–0.174, +0.321]	0.560	–0.039 (0.141)	0.782
Pre-election level (balance)	–0.031 [–0.144, +0.089]	0.648	–0.001 [–0.146, +0.172]	0.874	–0.157 [–0.419, +0.087]	0.199	+0.156 (0.152)	0.306
Municipal schools only (change in level)	–0.229 [–0.277, –0.187]	< 0.001	–0.308 [–0.384, –0.247]	< 0.001	–0.069 [–0.140, +0.023]	0.162	–0.238 (0.054)	< 0.001
State schools only (falsification)	–0.028 [–0.066, +0.009]	0.135	–0.037 [–0.094, +0.024]	0.244	+0.015 [–0.061, +0.101]	0.629	–0.052 (0.051)	0.308
Regime threshold: share appointed > 2/3	–0.199 [–0.259, –0.142]	< 0.001	–0.274 [–0.375, –0.195]	< 0.001	–0.116 [–0.201, +0.013]	0.026	–0.158 (0.066)	0.017
Regime threshold: share appointed > 1/3	–0.199 [–0.259, –0.142]	< 0.001	–0.253 [–0.337, –0.177]	< 0.001	–0.081 [–0.179, +0.037]	0.196	–0.172 (0.069)	0.012
Bandwidth × 0.5 to × 2 (range of estimates)	–0.226 to –0.193	all < 0.001	–0.313 to –0.250	all < 0.001	–0.147 to –0.080	0.006–0.184	—	—
Placebo cutoffs ±5 and ±10 pp (range)	–0.001 to +0.108	min 0.105	–0.012 to +0.073	min 0.275	—	—	—	—

Notes: Estimates with robust bias-corrected 95% confidence intervals and robust p-values. Specification as in Table 9; regime as in Table 10. INEP's teacher regularity indicator is available by school from 2016, so the 2012 cycle is excluded and the placebo is available only for 2020; post-election change = mean of the indicator in the two censuses after the election (May of years c+1 and c+2) minus its value in May of the election year. The standard deviation of the baseline outcome on the control side is 0.361.

were replaced by other temporary teachers and teachers were reassigned across schools, rather than if new hiring expanded the staff. For the 2016 and 2020 cycles, where both measures exist, the outcome pattern is the same as in the full sample: in appointment networks the harm index rises by 0.157 ($p = 0.10$) and grade 6–9 age-grade distortion by 3.58 points (CI 0.52 to 7.57).

7.6 Synthesis

The two political designs, read together, support predictions P3 to P5. Party turnover replaces municipal principals, including in elections decided by the narrowest margins, and the replacement is political in timing and location: it occurs immediately after the inauguration and in networks where the post is discretionary. On average across networks, close-election turnover has no detectable effect on learning or school flow, and the negative associations in the triple difference reflect the selection of municipalities whose schools were already deteriorating. Where the post is discretionary, however, turnover worsens grade progression in grades 6–9 and reshuffles the teaching staff, while estimates for networks with competitive selection are centered on zero.

The magnitudes are informative about mechanisms. When σ_θ is about 0.05, replacing a fifth of a network’s principals changes the network mean by far less than the confidence interval of any feasible design with public data, so the null on average achievement is what the variance estimates predict. The flow effects in appointment networks are too large to be explained by the replacement of principals alone. The selection regime is better read as a marker of networks in which the electoral calendar reaches deep into the school bureaucracy, and the teacher evidence supports this reading: in those networks party turnover lowers teacher regularity, while it leaves competitive networks and state schools essentially unaffected. Grade 6–9 progression is plausibly the margin most exposed to such disruption. It depends on subject teachers, who are more often hired on temporary contracts, accounting for 27–28% of municipal teachers in grades 6–9 in 2012–2020 against 22–23% in grades 1–5, and on promotion decisions taken within the school. Achievement in the tested grades moves less, or too little to be detected at the municipal level.

7.7 Relation to Akhtari, Moreira and Trucco (2022)

Our results are closely related to those of Akhtari, Moreira & Trucco (2022), who find that political turnover lowers test scores in Brazilian municipal schools by 0.05–0.08 standard deviations and attribute the decline to a broader turnover of school personnel. Our interval for grade-5 achievement (–0.10 to +0.04 SD) contains effects of that size, so the two sets of estimates are compatible, although our point estimate is smaller. The designs also differ in what they isolate. Ours compares municipal with state schools in the same municipality and therefore identifies the effect of turnover on the network the mayor controls, net of any municipality-wide consequences of the change of

government, and it pools four electoral cycles.

The variance estimates allow us to ask how much of an effect of that size could run through principals. Let f denote the share of principals replaced because of turnover and Δ the mean difference in effectiveness between induced successors and their predecessors, so that the effect of turnover on the network mean through this channel is $f\Delta$. With $f = 0.20$ from our first stage, a decline of 0.05–0.08 student SD requires Δ between -0.25 and -0.39 . If those estimates are expressed in standard deviations of school means, which equal about 0.44–0.51 student standard deviations in these data (Appendix Table A2), Δ must still lie between -0.11 and -0.20 . Gaps of this size are at least 1.6 times the upper bound on σ_θ (0.068) and at least 3.5 times the largest average change we measure at successions (0.031). A more plausible value for Δ is the gap we estimate between appointed and competitively selected successors, 0.026 SD, which yields $f\Delta \approx 0.005$ SD, between 7% and 11% of the decline in student units and between 13% and 24% in school units.

Principal replacement can therefore account for only a small part of the effect of political turnover on learning, and the remainder must operate through other personnel. Teachers are the natural candidates, and we show that their continuity falls sharply in the networks where principals are appointed. The regime pattern also provides a direct test of the interpretation proposed by Akhtari, Moreira & Trucco (2022), that turnover harms public services where the bureaucracy is not shielded from the political process. Measured by the rule used to select principals before each election, shielding matters: in networks that select principals competitively, the estimates for principals, teachers and school outcomes are all centered on zero.

8 Threats to Validity

8.1 The dating of successions

The variance estimator relies on tenure reported by principals in bands, and three features of these reports could bias it. First, recall error that shifts principals into an adjacent band compresses the differences in $\pi_{g,w}$ across windows and biases σ_θ^2 toward zero; in the simulation, misreporting by 15% of principals lowers the median estimate from 0.050 to 0.043. The stability of the estimates across values of q , and their agreement with and without group T, limit this concern. Second, arrivals may not be uniform within bands. If group-C arrivals were bunched after the January 2009 inaugurations, as the political first stage suggests for municipal networks, $\pi_{C, 2007-09}$ would exceed $2/3$ and σ_θ^2 would be overstated, but even under an extreme concentration with probability 0.90 the simulated median rises only from 0.050 to 0.053. Third, we observe only the principal in post in 2011. Very short spells between 2009 and 2011 followed by another change count as a single succession, which again biases the estimate downward, and if principal effects build up with tenure

our estimates weigh early-spell effects more heavily than long-run ones. When effects are halved during the first two years, a true long-run σ_θ of 0.075 produces a median estimate of 0.049, so a long-run dispersion of that size cannot be ruled out.

8.2 Selection on transitory shocks

The identifying assumption of Section 5.3 would fail if successions were triggered by unusually large shocks, since shock variances would then differ across groups in the succession window. Two pieces of evidence speak against this concern. The group-C covariance estimator, with 2007 as its base year, is immune to shocks in 2009 and 2011 and confirms the magnitude of the variance estimate. And in the simulation, selection of successions on negative shocks biases the variance estimator downward rather than upward.

8.3 Non-random selection regimes

Selection regimes are measured in 2011 with a leave-one-out share and are correlated with municipal characteristics. We have shown that the regime difference holds within municipal networks and outside the Northeast, survives interactions of the succession probability with rurality, socioeconomic status and region, and is not driven by any single state (Section 6.6), and the political first stage confirms that discretionary turnover occurs where the post is discretionary. These results make it unlikely that the regime simply proxies for disadvantage. They do not establish that changing a network’s selection rule would change σ_θ by the estimated difference, because networks that adopt competitive selection may differ in other, unobserved dimensions of administrative capacity, and we do not make that claim.

8.4 The political designs

Four features of the political designs deserve mention. The triple difference classifies schools by their 2007 dependency, and some state schools were later transferred to municipalities: 1.8% by 2011, 3.6% by 2015, 4.2% by 2019 and 4.7% by 2023. These transfers attenuate later-cycle estimates toward zero, which is why the regression discontinuity uses each school’s dependency in the year before each election. Election results come from CEM’s compilation of official results rather than directly from the electoral court, whose bulk files could not be retrieved programmatically; the compilation reproduces known aggregates, such as the 1,208 mayoralties won by the PMDB in 2008, and the alternation rates reported in Section 4.5. Pass rates jumped to about 98% in 2020–21 under automatic promotion policies adopted during the pandemic, so we report every result on political turnover with and without the 2020 cycle. Finally, the teacher regularity indicator is available only

from 2016, so the teacher analysis covers two cycles, one of them during the pandemic, and the pooled 2020 sample shows a pre-election imbalance that we report and net out.

8.5 Inference, multiple testing and specification search

The analysis was not pre-registered. We therefore distinguish primary from secondary estimands. The primary estimands are σ_θ by grade, its difference between regimes, the average effect of a succession and the regression-discontinuity effects on achievement. Heterogeneity by subject and by successor type, and the flow outcomes, are secondary, and the regression-discontinuity flow results are reported with Benjamini–Hochberg adjustment. For the two primary second-moment results we also report randomization inference, which does not rely on the asymptotic distribution of a clustered variance estimator and is more conservative for the regime difference.

Two analyses were motivated by earlier results. The regime split of the regression discontinuity follows from the hypothesis of Section 6.5 and from the first stage, but it was estimated after the pooled results were known. We test it with a single summary index and a municipal analogue of the classification in Section 6.5, and we report every alternative threshold and index definition we estimated, together with placebo cutoffs and pre-election balance. The effect within appointment networks is robust across these variants, whereas the difference between regimes is significant at the 5% level in only one of six, so we treat it as corroborating rather than independent evidence. The teacher analysis of Section 7.5 was designed to answer a specific question, namely whether the shock extends beyond principals, and we report it with its falsification test in state schools, its placebo and all the variants we estimated.

8.6 External validity

Successions are dated from a single cross-section in 2011, and the variance estimates refer to principals in post between 2005 and 2011. Several networks changed their selection rules after the 2014 national education plan and the 2020 Fundeb law, so the distribution of principals across regimes today differs from the one we observe. The political results cover four electoral cycles through 2020, and the teacher results the two most recent. Whether principal effects are equally small in systems where principals have more authority over staffing, as in many U.S. districts, is a question our design can inform only indirectly.

9 Discussion and Policy Implications

9.1 Why are principal effects small in this setting?

Our estimates place Brazil at the lower end of the range of principal effects reported in the literature, and it is worth asking why. Part of the answer is methodological. Estimates based on plug-in variances, few observations per spell and school-mean units can be several times larger than the underlying dispersion, and Section 6.4 shows that this sequence of choices alone reproduces the large estimates previously reported for Brazil. Once these choices are corrected, our grade-5 estimate of 0.05 is close to the recent U.S. evidence that little of the variation in test scores reflects persistent differences in principal effectiveness (Bartanen, Husain, and Liebowitz 2024) and below the estimates obtained with mobility designs in Texas and British Columbia (Branch et al. 2012; Dhuey and Smith 2014).

Part of the answer is likely institutional. Brazilian principals have limited formal authority over the recruitment and dismissal of teachers, who are hired by the network and assigned by the secretariat, whereas evidence from other settings links school productivity to how principals manage people (Lemos, Muralidharan, and Scur 2024). Principals also turn over quickly. Evidence from British Columbia indicates that principals take several years to reach their full effect on a school (Coelli and Green 2012), and with short spells our estimates weigh early-spell effects more heavily than long-run ones. The simulation evidence in Section 6.3 quantifies this concern: under effects that are halved during the first two years of a spell, the true long-run standard deviation consistent with our estimate would be about 0.075 rather than 0.05. Even this upper range implies modest gains from replacing individual principals. The grade-9 result fits the same account, since grade 9 is served mostly by state networks, which select their principals competitively and where P1 predicts little dispersion.

9.2 What the two halves of the paper imply together

The first half of the paper shows that the identity of the principal matters little for average achievement, and more where principals are appointed. The second half shows that political turnover replaces principals where they are appointed and that the resulting disruption harms grade progression and reaches the teaching staff. The link between the two is the degree of political discretion over school staff. In networks where the principalship is a political appointment, the distribution of management quality is wider, and the electoral calendar periodically reshuffles both principals and teachers. In networks with competitive selection, principals are closer to interchangeable and schools are insulated from the political cycle. The selection rule is thus better understood as a marker of the political governance of the school system than as a technology that affects outcomes

only through the quality of the principals it selects.

9.3 Implications for policy

The first implication concerns the expected return to changing principals. A standard deviation of principal-by-school effects of 0.05 student SD means that replacing a principal at the 16th percentile with one at the 84th raises school achievement by about 0.10 SD. That gain is meaningful but smaller than most estimates of the dispersion of teacher effects (Chetty, Friedman, and Rockoff 2014), and it agrees with a meta-analysis of 20 evaluations of school-management interventions in low- and middle-income countries, which estimates an average learning gain of 0.04 SD (Anand et al. 2023). Policies premised on principals explaining a large share of achievement differences, a premise that plug-in estimates of 0.2–0.9 appear to support, are likely to disappoint. Replacing principals in declining schools is no shortcut either, since successions follow deterioration and produce no average short-run gain.

The second implication concerns selection rules. Where principals are appointed, the dispersion of their effects is substantial, at about 0.08 SD in grade 5, whereas where they are selected competitively it is indistinguishable from zero, and competitively selected successors outperform appointed ones by 0.026 SD. These patterns fit a growing body of evidence on how public organizations recruit managers. In Chile, a civil-service reform that made principal selection more competitive and transparent led public schools to appoint more effective principals despite rigid wage schedules (Muñoz and Prem 2022). In Brazil, schools that moved from political appointment to election of principals improved measured management practices, and schools that moved in the opposite direction deteriorated by a similar amount (Cota, Flores, and Salgado 2025). Appointment is not uniformly harmful. A wider dispersion means that some appointed principals are highly effective, which is consistent with evidence that political connections can strengthen accountability and effort in the Brazilian bureaucracy under some conditions (Toral 2024b). The main benefit of merit-based selection is therefore protection against the lower tail of management quality rather than a higher average. The requirement of the national education plan and the Fundeb condition on principal selection should be judged by that metric: the relevant question is how often they prevent very weak appointments in the declining schools where successions concentrate, not how much they raise mean achievement.

The third implication concerns the political cycle. Mayoral turnover replaces principals only where the post is discretionary, and in close elections its costs appear where the post is discretionary, in the form of higher age-grade distortion and lower pass rates in grades 6–9. This complements evidence from the same institutional setting that political turnover lowers test scores in municipal schools through the turnover of school personnel (Akhtari, Moreira, and Trucco 2022), that out-

going governments reshuffle temporary staff and depress service delivery before the winner takes office (Toral 2024a), and that politicians move bureaucrats after changes in power in other federations (Iyer and Mani 2012). The policy options are institutional and inexpensive. They include fixed terms for principals that do not coincide with mayoral terms, restrictions on replacing school leadership during electoral transitions and selection procedures whose results bind the incoming administration. Because the reshuffling reaches teachers as well as principals, without changing the number of teachers or their mix of contracts, rules that protect principals alone would leave much of the disruption in place, and transition rules would need to cover the assignment of teachers to schools and the renewal of temporary contracts as well.

The fourth implication concerns investments in principals' skills. Experimental evidence shows that such investments are fragile when principals turn over. In Houston, 300 hours of management training raised achievement in the first year, but the second-year effect was insignificant, a result the author attributes to principal turnover (Fryer 2017). At scale, a comprehensive school-management program in India had no effect on school functioning or learning despite near-universal completion of its assessments (Muralidharan and Singh 2020), and the systematic review of Anand et al. (2023) identifies low take-up and weak incentives or structure for implementation among the factors that limit such programs. In Brazilian networks where the principalship changes hands with the mayor, the expected return to training is truncated by the electoral cycle. Training programs are therefore more likely to pay off where tenure is protected, and linking them to merit-based selection, as the Fundeb condition implicitly does, may matter as much as their content.

The fifth implication concerns measurement. Teacher ratings of leadership and principals' reports of practices predict growth before the principal's arrival as much as growth after it. Using such indices to evaluate or reward individual principals would largely reward the trajectories of the schools they inherited, and they are more useful as diagnostics of schools than as scores of their leaders.

9.4 Data infrastructure

Three additions to Brazil's public data would sharpen each of these results. A persistent, pseudonymized principal identifier in the School Census and SAEB would allow leave-out estimation of transferable principal effects and direct measurement of principal turnover in every electoral cycle. Recording the selection method every year would turn the regime classification from a 2011 snapshot into a panel and allow designs that exploit reforms induced by the Fundeb condition. A pseudonymized teacher identifier would allow direct measurement of transfers and contract renewals around elections, whereas the school-level regularity indicator used in Section 7.5 captures only their net effect. INEP's protected-data service already provides identified

access to the census and assessment data under its secure-room rules, and the designs in this paper could be replicated there with identifiers.

10 Conclusion

Public data are sufficient to identify the dispersion of principal-by-school effects in Brazil, provided one gives up the mobility of managers across schools and exploits instead the dated discontinuities of management within them. The standard deviation of these effects is about 0.05 student SD in grade 5 and statistically zero in grade 9, an order of magnitude below plug-in estimates computed from the same data. Changes of principal occur in declining schools and produce no average short-run gain.

The dispersion is concentrated in networks where principals are appointed, and competitively selected successors outperform appointed ones, as a simple model of selection on an index that mixes quality with political considerations predicts. Mayoral party turnover replaces municipal principals even in close elections. On average, close-election estimates find no detectable effect on learning or school flow, and the apparent harms estimated in panel designs reflect voters' responses to deteriorating schools. Where principals are appointed, however, turnover also displaces teachers and worsens grade progression in grades 6–9.

These findings shift the question that policy should ask. In this system, the identity of any individual principal matters less than the rules by which principals are chosen and the degree to which those rules expose schools to the political cycle. Reforms that make selection competitive and insulate school staff from electoral turnover are unlikely to raise average achievement by much through better principals alone, but they can protect the schools most at risk from very weak management and from the recurrent disruption of their staff.

Data Availability and Ethics

All data are public. We use SAEB/Prova Brasil 2011 microdata, school-level IDEB results (2005–2023), School Census microdata (2007–2024), school flow indicators (2007–2024), the teacher regularity indicator by school (2016–2024) and the statistical synopses of the School Census (2012–2022), all published by INEP, together with municipal election results from the Center for Metropolitan Studies (CEM-USP). No attempt was made to re-identify schools or principals in anonymized files. Replication code: `replicacao_sucessao_diretores.py` (variance estimators), `mc_sucessao.py` (Monte Carlo), `parse_fluxo_inep.py` (flow panel), `modulo_alternancia_cem.py` (elections, triple differences, stacked event studies),

`rd_empilhado.py` (stacked regression discontinuity), `canal_docentes.py` (teacher regularity and contract composition) and `make_figures_en.py` (figures).

References

- Akhtari, M., Moreira, D., & Trucco, L. (2022). Political turnover, bureaucratic turnover, and the quality of public services. *American Economic Review*, 112(2), 442–493. <https://doi.org/10.1257/aer.20171867>
- Anand, G., Atluri, A., Crawford, L., Pugatch, T., & Sheth, K. (2023). Improving school management in low and middle income countries: A systematic review. *Economics of Education Review*, 97, 102464. <https://doi.org/10.1016/j.econedurev.2023.102464>
- Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A*, 171(3), 673–697. <https://doi.org/10.1111/j.1467-985X.2007.00533.x>
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Bartanen, B., Grissom, J. A., & Rogers, L. K. (2019). The impacts of principal turnover. *Educational Evaluation and Policy Analysis*, 41(3), 350–374. <https://doi.org/10.3102/0162373719855044>
- Bartanen, B., Husain, A. N., & Liebowitz, D. D. (2024). Rethinking principal effects on student outcomes. *Journal of Public Economics*, 234, 105115. <https://doi.org/10.1016/j.jpubeco.2024.105115>
- Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic Journal*, 125(584), 647–674. <https://doi.org/10.1111/ecoj.12267>
- Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals. NBER Working Paper 17803. <https://doi.org/10.3386/w17803>
- Brazil (2014). Lei nº 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação.
- Brazil (2020). Lei nº 14.113, de 25 de dezembro de 2020. Regulamenta o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb).
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326. <https://doi.org/10.3982/ECTA11757>
- Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531), 1449–1455. <https://doi.org/10.1080/01621459.2019.1635480>
- Cauduro, M. A. S. (2018). *Ensaio em liderança e gestão* [Essays on leadership and management]. Doctoral dissertation, Fundação Getulio Vargas, São Paulo School of Business Administration.
- Centro de Estudos da Metrópole (n.d.). *Resultados eleitorais por município no Brasil, 2004–2020* [Electoral results by municipality]. University of São Paulo. centrodametropole.fflch.usp.br
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632. <https://doi.org/10.1257/aer.104.9.2593>
- Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 31(1), 92–109. <https://doi.org/10.1016/j.econedurev.2011.09>

- Colonnelli, E., Prem, M., & Teso, E. (2020). Patronage and selection in public sector organizations. *American Economic Review*, 110(10), 3071–3099. <https://doi.org/10.1257/aer.20181491>
- Cota, G., Flores, A., & Salgado, M. (2025). Voting for managers: Selection methods in public schools. SSRN Working Paper 5282703. <https://ssrn.com/abstract=5282703>
- Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics*, 47(2), 634–663. <https://doi.org/10.1111/caje.12086>
- Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica*, 90(3), 1063–1084. <https://doi.org/10.3982/ECTA19244>
- Ferraz, C., & Finan, F. (2008). Exposing corrupt politicians: The effects of Brazil’s publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2), 703–745. <https://doi.org/10.1162/qjec.2008.123.2.703>
- Fryer, R. G., Jr. (2017). Management and student achievement: Evidence from a randomized field experiment. NBER Working Paper 23437. <https://doi.org/10.3386/w23437>
- Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How Principals Affect Students and Schools: A Systematic Synthesis of Two Decades of Research*. The Wallace Foundation. <https://doi.org/10.59656/el-sb1065.001>
- Iyer, L., & Mani, A. (2012). Traveling agents: Political change and bureaucratic turnover in India. *Review of Economics and Statistics*, 94(3), 723–739. https://doi.org/10.1162/REST_a_00183
- Kline, P., Saggio, R., & Sølvssten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5), 1859–1898. <https://doi.org/10.3982/ECTA16410>
- Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1), 83–119. <https://doi.org/10.1111/j.1468-0262.2007.00733.x>
- Lemos, R., Muralidharan, K., & Scur, D. (2024). Personnel management and school productivity: Evidence from India. *The Economic Journal*, 134(661), 2071–2100. <https://doi.org/10.1093/ej/uead112>
- Muñoz, P., & Prem, M. (2022). Managers’ productivity and recruitment in the public sector: The case of school principals. TSE Working Paper 22-1303, Toulouse School of Economics.
- Muralidharan, K., & Singh, A. (2020). Improving public sector management at scale? Experimental evidence on school governance in India. NBER Working Paper 28129. <https://doi.org/10.3386/w28129>
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5), 2555–2591. <https://doi.org/10.1093/restud/rdad018>
- Tavares, P. A. (2015). The impact of school management practices on educational performance: Evidence from public schools in São Paulo. *Economics of Education Review*, 48, 1–15. <https://doi.org/10.1016/j.econedurev.2015.05.002>
- Toral, G. (2024a). Turnover: How lame-duck governments disrupt the bureaucracy and service delivery before leaving office. *Journal of Politics*, 86(4), 1348–1367. <https://doi.org/10.1086/729961>
- Toral, G. (2024b). How patronage delivers: Political appointments, bureaucratic accountability, and

service delivery in Brazil. *American Journal of Political Science*, 68(2), 797–815. <https://doi.org/10.1111/ajps.12758>

Online Appendix

Table A1—Mean squared residual of two-year changes by group and window (grade 5, mathematics)

Group	2005–07	2007–09	2009–11	2011–13
S (stable)	0.0975	0.1160	0.1154	0.1122
D (arrival 2004–06)	0.0860	0.1126	0.1142	0.1089
C (arrival 2006–09)	0.0966	0.1169	0.1078	0.1186
T (succession 2009–11)	0.0849	0.1038	0.1047	0.1105

Bold: window containing the succession. Values do not control for sampling noise; Table 4 uses the version with the noise control and group fixed effects.

Table A2—Plug-in versus noise-corrected estimates (student SD)

	Grade 5	Grade 9
Succession / stable schools	6,974 / 4,800	4,751 / 3,540
Plug-in SD of the difference between spells	0.249	0.227
Expected noise SD of that difference	0.248	0.221
Noise-corrected SD of the difference	0.020	0.053
Plug-in σ_θ (student SD)	0.176	0.161
Plug-in σ_θ (school-mean SD)	0.344	0.365
Noise-corrected σ_θ	0.014	0.037

Table A3—Do leadership and practice indices measured in 2011 predict achievement changes? (grade 5)

Index (z)	Change	Stable schools: slope (s.e.)	Interaction with succession (s.e.)	Schools
Teacher-rated leadership	2005→09 (placebo)	0.030 (0.005)	−0.013 (0.006)	11,848
Teacher-rated leadership	2009→11	0.009 (0.005)	−0.005 (0.006)	11,848
Teacher-rated leadership	2009→13	0.014 (0.005)	−0.003 (0.007)	10,577
Teacher-rated leadership	2009→15	0.002 (0.006)	+0.006 (0.007)	10,231
Principal-reported practices	2005→09 (placebo)	0.014 (0.005)	+0.001 (0.007)	11,849
Principal-reported practices	2009→11	0.008 (0.005)	−0.006 (0.006)	11,849
Principal-reported practices	2009→13	0.006 (0.006)	+0.006 (0.007)	10,579
Principal-reported practices	2009→15	0.005 (0.006)	+0.005 (0.007)	10,235

Notes: State × network × location fixed effects and school SES; errors clustered by municipality. If the indices measured the causal effect of the current principal, slopes in stable schools and in the placebo window would be zero, and the interaction in post-arrival windows would be positive.

Table A4—Robustness of the stacked regression discontinuity (estimate [robust 95% CI]; p)

Specification	Achievement, grade 5	Achievement, grade 9	Age-grade distortion, 6–9	Repetition, 1–5
Baseline (MSE bandwidth, local linear)	−0.021 [−0.100, +0.044]; 0.45	+0.000 [−0.068, +0.070]; 0.97	+1.39 [+0.13, +2.86]; 0.03	−1.32 [−2.64, −0.29]; 0.01
Without 2020	−0.013 [−0.103, +0.065]; 0.66	−0.011 [−0.081, +0.056]; 0.73	+0.88 [−0.68, +2.59]; 0.25	−1.34 [−2.88, −0.19]; 0.02
Half bandwidth	−0.035 [−0.132, +0.112]; 0.87	−0.001 [−0.098, +0.139]; 0.73	+2.25 [−0.03, +4.77]; 0.05	−1.25 [−3.39, +0.62]; 0.18
Double bandwidth	−0.007 [−0.082, +0.052]; 0.66	−0.003 [−0.064, +0.062]; 0.99	+1.23 [+0.01, +2.55]; 0.05	−1.05 [−2.65, −0.49]; 0.00
Local quadratic	−0.037 [−0.134, +0.052]; 0.38	+0.001 [−0.083, +0.085]; 0.98	+2.60 [+0.95, +4.79]; 0.00	−1.58 [−2.88, −0.46]; 0.01
Donut (exclude margins below 0.5 pp)	−0.014 [−0.098, +0.061]; 0.65	−0.012 [−0.084, +0.058]; 0.72	+1.17 [−0.30, +2.66]; 0.12	−1.19 [−2.87, +0.11]; 0.07
2008 only	−0.047 [−0.168, +0.072]; 0.43	+0.032 [−0.077, +0.166]; 0.48	+1.07 [−1.56, +4.06]; 0.38	−2.61 [−5.32, −0.53]; 0.02
2012 only	−0.003 [−0.170, +0.133]; 0.81	−0.026 [−0.147, +0.080]; 0.56	+1.69 [−0.65, +4.73]; 0.14	−0.87 [−2.59, +0.80]; 0.30
2016 only	+0.015 [−0.116, +0.155]; 0.78	−0.016 [−0.182, +0.130]; 0.75	+0.80 [−3.17, +4.49]; 0.74	+0.38 [−1.34, +1.94]; 0.72
2020 only	−0.085 [−0.277, +0.065]; 0.22	+0.092 [−0.147, +0.373]; 0.39	+4.67 [+2.02, +8.46]; 0.00	−1.60 [−3.62, +0.18]; 0.08
Post minus pre-election change (2012–2020)			+1.44 [−0.81, +3.75]; 0.20	−0.54 [−2.53, +1.21]; 0.49

Table A5—Monte Carlo: estimator performance under the design assumptions and under violations (grade 5)

Scenario	Median $\hat{\sigma}_\theta$, true 0.05	Median $\hat{\sigma}_\theta$, true 0.10	Coverage of 95% CI, true 0.05	Rejection of $\sigma_\theta = 0$, true 0	Power, true 0.05
Design assumptions hold	0.050	0.100	0.947	0.019	0.829
Shocks with $t(5)$ tails	0.050	0.100	0.972	0.025	0.742
$q = 0.15$ (estimator uses 0.33)	0.052	0.106	0.930	0.022	0.898
$q = 0.50$ (estimator uses 0.33)	0.047	0.093	0.918	0.028	0.722
Arrivals bunched after Jan 2009	0.053	0.106	0.920	0.035	0.932
15% report adjacent tenure band	0.043	0.087	0.880	0.010	0.602
Half effect in first two years	0.034	0.067	0.618	0.032	0.260
Succession after negative shock	0.025	0.089	0.310	0.000	0.108

Notes: 1,000 replications per value under the design assumptions and 400 under each violation; coverage refers to $2\sigma_\theta^2$; rejection is one-sided at 2.5%. Coverage under violations is computed against the true simulated parameter.

Table A6—Placebo cutoffs in the stacked regression discontinuity (estimate; robust p in parentheses)

Outcome	Cutoff −10 pp	Cutoff −5 pp	Cutoff +5 pp	Cutoff +10 pp
Achievement, grade 5 (student SD)	−0.116 (0.16)	−0.009 (0.69)	−0.069 (0.19)	+0.044 (0.47)
Achievement, grade 9 (student SD)	+0.052 (0.49)	+0.033 (0.67)	−0.042 (0.66)	−0.028 (0.75)
Age-grade distortion, grades 6–9 (pp)	−0.37 (0.69)	−1.13 (0.40)	+2.05 (0.20)	+1.55 (0.56)
Pass rate, grades 6–9 (pp)	−0.70 (0.67)	+2.96 (0.048)	−0.37 (0.87)	−0.78 (0.63)
Repetition, grades 1–5 (pp)	+1.02 (0.34)	−1.27 (0.21)	−0.70 (0.67)	−0.75 (0.67)
Harm index, appointment networks, 2012–2020	−0.054 (0.64)	−0.213 (0.18)	+0.062 (0.55)	+0.003 (0.95)
Harm index, competitive or mixed networks, 2012–2020	−0.249 (0.07)	+0.109 (0.68)	+0.201 (0.09)	+0.161 (0.46)

Notes: Each placebo uses only observations on one side of the true cutoff (losers of the incumbent party for positive cutoffs, winners for negative ones), recentred at the false cutoff. Specification as in Table 9. For the five individual outcomes, 19 of 20 placebo estimates have $p > 0.05$; the exception is the grade 6–9 pass rate at −5 points ($p = 0.048$). No placebo estimate for the harm index is significant at 5% in either regime.

Table A7—Components of the harm index by selection regime, close elections 2012–2020
(estimate [robust 95% CI])

Outcome (municipal – state change)	Appointment networks	p	Competitive or mixed networks	p
Achievement, grade 5 (student SD)	–0.049 [–0.201, +0.055]	0.261	+0.002 [–0.108, +0.133]	0.836
Achievement, grade 9 (student SD)	–0.043 [–0.150, +0.053]	0.347	+0.027 [–0.107, +0.163]	0.683
Age-grade distortion, grades 6–9 (pp)	+2.98 [+0.84, +5.99]	0.009	+1.71 [–0.97, +4.10]	0.227
Pass rate, grades 6–9 (pp)	–2.10 [–5.42, +0.66]	0.125	+1.19 [–1.98, +4.07]	0.499
Repetition, grades 6–9 (pp)	+1.44 [–1.02, +4.24]	0.231	–0.22 [–2.64, +2.56]	0.976
Dropout, grades 6–9 (pp)	+0.80 [–0.47, +2.50]	0.179	–1.04 [–3.21, +0.81]	0.243
Age-grade distortion, grades 1–5 (pp)	+0.54 [–2.30, +3.68]	0.651	–0.22 [–3.84, +3.30]	0.882
Pass rate, grades 1–5 (pp)	–0.23 [–2.39, +2.20]	0.935	+0.56 [–2.09, +2.97]	0.733
Repetition, grades 1–5 (pp)	–0.57 [–2.87, +1.47]	0.528	–1.66 [–3.84, +0.14]	0.069
Dropout, grades 1–5 (pp)	+0.81 [–0.12, +1.85]	0.086	+1.64 [–0.48, +4.64]	0.112

Notes: Regime and specification as in Table 10. Individual components are reported without multiplicity adjustment; the summary index in Table 10 provides a single test of the joint pattern.

Table A8—Party turnover in close elections and the contract composition of teachers

Teachers in grades 1–9 (Sinopse Estatística; shares 0–1 and logs)

Outcome	All	p	Appointment	p	Competitive or mixed	p	Difference (s.e.)	p
Temporary share, municipal – state, 2012–2020								
Without the 2020 cycle	+0.010 [–0.021, +0.034]	0.656	+0.001 [–0.055, +0.043]	0.819	+0.002 [–0.049, +0.047]	0.974	–0.001 (0.035)	0.988
Placebo: pre-election change (2016, 2020)	+0.027 [–0.005, +0.054]	0.100	+0.005 [–0.058, +0.057]	0.988	–0.002 [–0.061, +0.050]	0.845	+0.007 (0.041)	0.857
Pre-election level (balance)	+0.006 [–0.038, +0.059]	0.676	+0.013 [–0.073, +0.100]	0.767	–0.038 [–0.132, +0.044]	0.327	+0.051 (0.063)	0.419
Temporary share, municipal network only	+0.023 [–0.016, +0.073]	0.215	+0.029 [–0.038, +0.104]	0.361	+0.047 [–0.034, +0.141]	0.232	–0.018 (0.057)	0.757
Temporary share, state network (falsification)	+0.016 [–0.002, +0.031]	0.081	+0.010 [–0.015, +0.033]	0.462	+0.005 [–0.019, +0.036]	0.558	+0.005 (0.019)	0.792
log number of municipal temporary teachers	–0.003 [–0.019, +0.017]	0.910	–0.002 [–0.035, +0.040]	0.889	+0.003 [–0.032, +0.045]	0.732	–0.004 (0.027)	0.872
log number of municipal permanent teachers	+0.017 [–0.094, +0.099]	0.960	+0.090 [–0.085, +0.232]	0.363	–0.154 [–0.454, +0.082]	0.173	+0.244 (0.159)	0.125
log number of municipal teachers, total	–0.031 [–0.059, –0.007]	0.014	–0.028 [–0.065, +0.010]	0.150	+0.003 [–0.051, +0.042]	0.845	–0.030 (0.030)	0.317
	+0.007 [–0.012, +0.022]	0.574	–0.002 [–0.026, +0.015]	0.599	+0.016 [–0.015, +0.042]	0.341	–0.019 (0.018)	0.288

Notes: Estimates with robust bias-corrected 95% confidence intervals and robust p-values. Specification as in Table 9; regime as in Table 10. Data are municipality-level counts of teachers in the public network of grades 1–5 and 6–9 by employment status (permanent civil servant, temporary contract, outsourced, CLT) and network, INEP Sinopse Estatística 2012–2022, tables 2.24 and 2.28; shares require at least five teachers; 2012 municipalities matched by state and name (seven unmatched).