Accountability Backlash: Negative Electoral Responses to Public Service Provision in Brazil

July 5, 2019

Abstract

Do voters reward politicians for the quality of public services? We address this question by studying voters' responses to signals of municipal school quality in Brazil, a setting particularly favorable to electoral accountability. Findings from two research designs—a regression discontinuity and a field experiment—are strikingly consistent. Contrary to expectations, signals of school quality *decrease* electoral support for the local incumbent. However, we find the expected effect among citizens for whom school quality should be most salient—parents with children in municipal schools. Using an online survey experiment to test the mechanism behind this result, we argue that voters interpret educational quality as an indicator of municipal spending priorities and perceive trade-offs with other areas that they might value more. We propose the concept of "accountability backlash" to make sense of our findings as well as of a growing body of evidence of negative voter responses to service delivery.

1 Introduction

Do voters hold governments accountable for the delivery and quality of public services? Governments and NGOs around the world are increasingly embracing transparency initiatives to foster electoral accountability (Gaventa and McGee, 2013; Barber et al., 2015). The key idea behind these initiatives is that voters have information constraints that prevent them from holding elected officials accountable for the quality of public services. Therefore, the logic goes, providing relevant and timely information will enable and empower them to punish bad performers and reward good ones, thus inducing the selection of better politicians and giving them more incentives to perform.

While the accountability logic is powerful, recent experimental evidence from around the world points to the limits of performance-based accountability systems. First, a number of studies have found that voters fail to hold politicians accountable even when provided with relevant information in a timely manner (Boas and Hidalgo, 2019; Dunning et al., 2019,?). Second, some research shows that information prompts accountability voting only for (sometimes small) subgroups of the population or under a specific set of conditions, and not for the electorate in general (Adida et al., 2017,?; Boas et al., 2019). Third, and most worrisome from a normative standpoint, several studies suggest that electoral accountability sometimes works in unexpected and undesirable directions, with voters punishing good performers and rewarding bad ones (Adida et al., 2017; Arias et al., 2018; Blattman et al., 2018; Chong et al., 2015; De Kadt and Lieberman, 2017). With transparency initiatives increasingly common around the world, it has become urgent to understand under what conditions and for what kind of voters information can foster electoral accountability.

To address these questions, we study how information about municipal school quality impacts voting behavior and electoral outcomes for local incumbents in Brazil. The Brazilian education system is a favorable context for voters to hold municipal politicians accountable for the quality of public schools. Municipalities are responsible for primary

education, and they are expected to spend at least a quarter of their revenue on schools. The federal government regularly releases a simple, credible, and highly visible metric of the quality of municipal schools, and citizens receive this binary signal of school quality shortly before elections. Moreover, media and citizens appear to pay significant attention to such signals. These conditions make it plausible that citizens respond to school quality signals by supporting good performers or voting against bad ones.

Empirically, we approach our research question with two complementary research designs. First, we employ a regression discontinuity design to compare municipalities across Brazil that barely met their school quality target and those that barely missed it. Second, we use a field experiment providing information about school quality to a random sample of voters in municipalities in the state of Pernambuco. This unique combination of designs allows us to examine both the macro- and the micro-level effects of school quality signals on electoral outcomes. While the quasi-experimental study allows us to measure effects in a naturally occurring environment, thus addressing issues of external validity, the experimental study allows us to assuage potential concerns about the internal validity of regression discontinuity models and to more finely test for heterogeneous treatment effects.

The results from the regression discontinuity analysis and field experiment are strikingly consistent: positive signals of school quality are found to *decrease*, rather than increase, support for the incumbent.[1] Municipalities that meet their school quality target see the electoral performance of the mayor decrease, and individuals who are informed about positive school performance in their municipality are less likely to vote for the incumbent. In the case of the field experiment, these findings contradict pre-registered hypotheses. Yet we also uncover significant heterogeneity in individual-level treatment

---

[1]These research designs were not coordinated a priori; we independently arrived at similar, counterintuitive conclusions in both the RDD and field experiment and decided to combine our results in a single write-up.

effects. In line with the pre-analysis plan for the experiment, parents of children enrolled in municipal schools respond to information about school quality in the expected direction, punishing bad performers rather than good ones.

To test the psychological mechanism underlying our findings, we conducted an online survey experiment in which we randomly provided information about local school quality to a diverse sample of Brazilian voters. Though there is no empirical relationship between municipal education spending and measures of school quality, we find that voters perceive such a relationship: when given a positive (negative) signal, respondents were more (less) likely to agree that the mayor invested a lot in education. Combined with recent experimental evidence that poor voters in Brazil disapprove of increased educational spending because they prefer conditional cash transfers (Bursztyn, 2016), these results suggest that trade-off thinking underlies our counterintuitive results. Voters interpret school quality signals as evidence of the mayor's spending priorities, and the majority who do not have a direct stake in this policy area punish good performance because they prefer that the government spend its money elsewhere.

We argue that our findings constitute evidence of an "accountability backlash" in which voters punish the delivery of desirable policy outputs and reward the delivery of undesirable ones. In this respect, our results accord with those of several other experimental or quasi-experimental studies that have reached similar conclusions. Our findings suggest that theories of electoral accountability and retrospective voting need to account for potential negative voter responses to the delivery of public services. On the policy front, our results should serve as a warning about the potential unintended consequences of performance-based accountability systems.

2 Information, Accountability, and Accountability Backlash

Voters around the world often lack information about the performance of their governments, and poor-performing politicians and parties are routinely returned to office. These twin facts have given rise to the hypothesis that providing voters with timely and relevant information about government performance will allow them to take action at the polls, voting against poor performing parties and politicians and in favor of those that govern effectively (Dunning et al., 2019). Inspired by this notion, a number of governments and NGOs have developed initiatives seeking to increase citizens' access to information about government performance.

An initial body of evidence suggested that information and accountability initiatives can work as intended (Pande, 2011). In Brazil, Ferraz and Finan (2008) show that random audits of municipal governments reduced incumbents' reelection prospects when negative information was uncovered. In India, Banerjee et al. (2011) found that performance information boosted the vote share of better performing and more qualified incumbents.

Yet subsequent research has cast doubt upon the information and accountability hypothesis. First, a number of studies suggest that even relevant and timely information about incumbent performance can have null effects on voting behavior. In a Brazilian mayoral election, flyers conveying corruption allegations against each candidate in the runoff reduced vote share only for the challenger, not for the incumbent (De Figueiredo et al., 2011). In Uganda, delivering information about incumbent legislators' performance had no effect on their vote shares or reelection prospects (Humphreys and Weinstein, 2012). Most recently, six separate studies in Africa and Latin America, informing about aspects of performance from public goods provision to charges of malfeasance, show that information provision almost always has null effects on voting for the incumbent (Dunning et al., 2019,?). Though voters often react strongly to information on incumbent performance in the context of hypothetical vignettes, they may fail to act on the same information when delivered in real life (Boas et al., 2019; Weitz-Shapiro and Winters, 2017).

Second, a number of recent studies demonstrate that delivering information to voters can prompt electoral accountability, but only in particular subgroups of the population or under a unique set of circumstances. In a field experiment in Benin, Adida et al. (2017) find that ethnicity moderates the effect of information on electoral accountability: voters reward good performers only if they are co-ethnics and punish bad performers only if they are non-co-ethnics. In the same experiment, information could also prompt accountability voting more generally, but only when widely disseminated to facilitate coordination and also combined with a "civics message" that reinforced the salience of the information itself (Adida et al., 2017). In a field experiment in Uganda, Buntaine et al. (2018) find that voters who receive information about local government irregularities punish bad performers only when they are running for lower-level positions. And in Brazil, Boas et al. (2019) show that negative information about local governments' mosquito control efforts prompts voting against the incumbent only for respondents who know someone with a child affected by the Zika virus.

Third, a number of studies have found that information and accountability systems can backfire, with voters punishing incumbents for good performance or rewarding them for bad performance. In Benin, informing voters about good performance by their legislator (such as attending and speaking at legislative session) prompted a punishment effect because voters assumed a trade-off with particularistic transfers, which they valued more (Adida et al., 2017). Arias et al. (2018) find that detailed revelations of wrongdoing by mayors in Mexico increased support for the mayor's party because many voters had uncertain or highly negative prior beliefs about their levels of malfeasance. In an earlier study of Mexico that distributed a similar set of audit reports, Chong et al. (2015) find that information about incumbent malfeasance had a demobilizing effect that worked to the net benefit of the incumbent party.

There is also evidence that real-world service delivery can have unexpected backlash effects. De Kadt and Lieberman (2017) argue that improved service provision lowers

support for the incumbent in several Southern African democracies because it increases exposure to corruption and raises voters' expectations of government performance. In Uganda, Blattman et al. (2018) show that a lottery-based program providing cash grants to poor entrepreneurs increased support for the opposition party because it raised recipients' incomes and freed them from reliance on patronage networks. These outcomes are less troubling from a normative standpoint—punishing corruption, higher expectations of government, and freedom from patronage networks are all positive consequences— but political backlash effects from the introduction of good programmatic policies could undercut parties' incentives to provide them in the future.

Recent experimental studies are expanding our understanding of how and when information leads to electoral accountability. Yet while they allow for rigorous identification of causal effects and their heterogeneity across individuals, they often raise questions of external validity and general equilibrium effects. Our unique combination of a field experiment and a natural experiment allows us to approach the question from micro- and macro-level perspectives, in a setting where, as the next section shows, institutions are ex ante conducive to a link between information and accountability.

Substantively, we argue that accountability backlash can arise because voters perceive trade-offs not only between programmatic and particularistic performance (Adida et al., 2017) but also between different policy priorities. We show that parents of children enrolled in local schools respond to school quality signals by punishing poor performing mayors, as expected, but that other voters respond in the opposite direction, punishing good performance. We argue that voters interpret school quality signals as evidence of the mayor's spending priorities, and that those without a direct stake in municipal public education may prefer that scarce resources be spent elsewhere. Our analysis thus complements that of Bursztyn (2016), who demonstrate a trade-off with respect to spending on public education versus direct cash transfers. Here, we show that information about policy outputs can also prompt trade-off thinking if voters infer that improved public

6

service provision implies a reallocation of budget resources.

3   Institutional Setting

Brazilian municipal education is a unique setting in which to study electoral responses to the quality of public services, since the country has a high-quality, high-visibility system for measuring the quality of public education, and education is one of the most important policy areas for municipal governments. Brazil has 5,570 municipalities, with municipal elections held every four years in October. Mayors are elected by simple plurality vote in the vast majority of municipalities and by two-round majority runoff in municipalities with over 200,000 inhabitants, and they are limited to two consecutive terms. Once elected, municipal governments are required to spend at least a quarter of their revenue on local education.

Brazilian basic education is structured in two cycles: primary school (grades 1 through 5) and middle school (grades 6 through 9). There are private schools at both levels, but most families opt for the public school system, which enrolled more than 81% of primary school students in Brazil's 2018 School Census. Public schools can be managed by any level of government, but municipalities are mostly responsible for primary education (83% of public school enrollments in 2018), while state governments usually run middle and high schools.

While basic education is mostly in hands of sub-national governmenled *Prova Brasil*), which tests students in fifth and ninth grades (i.e., at the end of primary and middle school). ANA is implemented every year, and ANRESC is implemented every two years. Exams are based on item response theory, which ensures that its measures of learning outcomes are valid and comparable over time.

After ANRESC was first implemented in 2005, the federal government created the Basic Education Development Index (IDEB, *Índice de Desenvolvimento da Educação Básica*)

to measure and incentivize educational performance. IDEB multiplies average ANRESC test scores by passing rates, so as to avoid perverse incentives for schools to either automatically pass children or hold them back to boost test scores. The government established IDEB targets for the country as well as all schools, municipalities, and states for every two-year period from 2007 to 2021. These targets were defined based on an algorithm that projected school progress along logistic trajectories with the goal of getting the country to average performance levels in OECD countries by 2021 (Fernandes, 2007). Targets took into account baseline levels of performance and were thus lower for initially weaker schools, municipalities, and states. Once released at the beginning of the period, IDEB targets have not been revised. World Bank economists have called IDEB "one of the world's most impressive systems for measuring education results [...], superior to current practice in the United States and in many other OECD countries in the quantity, relevance, and quality of the student and school performance information it provides" (Bruns et al., 2012, 7).

By providing an easy-to-understand, binary performance metric—whether or not the target was met—IDEB results are particularly visible and influential. Brazilian media pay significant attention to IDEB scores and whether targets are met, especially during the days immediately after the federal government releases the results.[2] As a newer test that does not involve targets, ANA is somewhat less visible than IDEB and ANRESC, though it also attracts media attention and makes headlines after results are released.

There is also evidence of citizen demand for indicators of school performance, at least among those who care about education. As shown in Appendix A2, Google searches for "IDEB" are very common after results are released, even compared to other performance-

---

[2]For example, *Folha de São Paulo*, the largest newspaper in the country, has always run a cover story on IDEB the day after the government publishes the results. Small media and blogs also pay attention to IDEB, generally focusing on whether targets were met. See Appendix A1 for some quantitative evidence of media coverage of IDEB.

related terms such as corruption, inflation, and the conditional cash transfer program Bolsa Família. In our representative, face-to-face survey of voters in the state of Pernambuco (details below), we found that high test scores were the second most cited quality of a good school (21% of respondents), after having well-trained teachers (33%).

The schedule of the release of IDEB scores further facilitates electoral accountability. In recent years, IDEB results have been made public about a month before elections are held (see Appendix A3). This timing ensures that results are in the public eye at a time when the media and citizens are evaluating government performance, incumbents are claiming accomplishments, and challengers are highlighting their shortcomings.

Given municipal responsibility for education, the existence of clear performance metrics, and evidence of citizen interest in the information, it is reasonable to expect that informing citizens about educational performance prompts electoral accountability. Providing individual voters with information about standardized test scores should lead them to reward good performance by voting for the mayor's reelection and punish bad performance by voting against it. Moreover, given the visibility and easy-to-understand nature of IDEB results and the timing of their release, it is reasonable to expect that meeting versus missing the IDEB target prompts electoral accountability on its own, without the need for an outside intervention.

4   Research Designs

To test these hypotheses regarding educational performance information and electoral accountability, we rely on two different research designs. First, we use a regression discontinuity design to identify the effect of meeting the IDEB target on electoral outcomes in municipalities across Brazil. Second, we analyze a field experiment in the state of Pernambuco that randomly assigned citizens to receive information about their municipality's performance on the ANA, examining the effect of this treatment on vote for the

9

incumbent mayor's reelection. We thus combine two different empirical strategies and measures of education quality to study whether and how voters respond to signals of public education quality.

4.1 Design 1: Regression Discontinuity

Regression discontinuity designs (RDDs) examine the effect of a treatment that is assigned deterministically by surpassing an arbitrary threshold of an underlying continuous variable. In the present case, the difference between the IDEB score and the IDEB target for a given municipality gives us a continuous measure of its performance. If that difference is zero or greater, the municipality met or surpassed its target and receives the treatment; if it is negative, the municipality missed its target and is in the control condition. Subject to assumptions discussed below, this design allows us to interpret a discontinuous jump of the outcome variable at the threshold as the causal effect of the treatment, namely meeting the IDEB target.

More formally, the treatment status for municipality m in period j, T_{mj}, is assigned by the forcing variable, which is the difference between that municipality's IDEB score and IDEB target ($D_{mj} = score_{mj} - target_{mj}$). While the Ministry of Education uses figures with one decimal only, we use a continuous measure to increase statistical power and avoid the issues with discrete forcing variables in RDDs (Lee and Card, 2008). The cutoff is therefore at –0.05 in the continuous measure, which is equivalent to 0 with the rounding applied by the Ministry:

$$
T_{mj} = \begin{cases} 1 & \text{if } D_{mj} \geq -0.05 \quad \text{(rounding, IDEB score} \geq \text{IDEB target)} \\ 0 & \text{if } D_{mj} < -0.05 \quad \text{(rounding, IDEB score} < \text{IDEB target)} \end{cases} \tag{1}
$$

Our estimand of interest is $\tau = \mathbb{E}[Y_{1i,j} - Y_{0i,j}]$, where $Y_{1i,j}$ and $Y_{0i,j}$ represent the potential outcome of interest (vote share or re-election of the mayor) under treatment (having

10

met the IDEB target) and under control (having missed it). Note that we only observe the outcome of a given municipality either under treatment or under control. That is to say, $Y_{1i,j}|T_{i,j} = 0$ and $Y_{0i,j}|T_{i,j} = 1$ are unobserved: this is the fundamental problem of causal inference. As long as average potential outcomes are continuous, we can estimate the local average treatment effect (LATE) around the cutoff $c = -0.05$ by taking the difference in means from above and below the threshold:

$$\tau = \mathbb{E}[Y_{1mj} - Y_{0mj}|D_{mj} = c] = \lim_{D_{mj}\downarrow c} \mathbb{E}[Y_{1mj}|D_{mj} = c] - \lim_{D_{mj}\uparrow c} \mathbb{E}[Y_{0mj}|D_{mj} = c] \quad (2)$$

This is the LATE for municipalities around the threshold, namely with scores slightly below and slightly above their targets. Since we are interested in the effect of meeting the target, the LATE for units close to the threshold (i.e., those that may plausibly switch from treatment to control, or vice-versa) is a meaningful quantity of interest.

The key assumption of this design is that potential outcomes are continuous around the threshold, so that the mean of the outcome of municipalities barely treated is a valid counterfactual for the mean of the outcome of municipalities barely untreated. Formally, we are assuming that $\mathbb{E}[Y_{dmj}|D_{mj} = d]$ is continuous in d around $D_{mj} = -0.05$ for both the treatment and the control groups (Imbens and Lemieux, 2008). While this assumption is empirically untestable, we can examine some of its observable implications. A key implication is that municipalities do not sort around the threshold. If we observed that municipalities cluster on the right-hand side of the threshold, we could suspect that local governments are manipulating their scores in order to reach their targets. Appendix A4 shows the forcing variable has a roughly normal distribution with no signs of sorting around the threshold, and the null hypothesis of continuity around the threshold cannot be rejected using the test proposed by McCrary (2008). Appendix A5 shows there are no discontinuous jumps in pre-treatment covariates either.

4.1.1 Data

For election outcomes, we use data from Brazil's Superior Electoral Court. For IDEB scores and targets, we use the Ministry of Education's IDEB results for primary education at the level of the municipality. For balance checks and further specifications, we use data from the 2010 census and from the "Basic Municipal Information" dataset for 2009, both administered by Brazil's official statistics agency (IBGE, *Instituto Brasileiro de Geografia e Estatística*), as well from the Ministry of Education's yearly school census. We use three IDEB waves (2007, 2011 and 2015), the results of which were published before the municipal elections of 2008, 2012 and 2016. Our effective sample excludes municipality-period observations where the mayor is not eligible to run for reelection due to term limits. When using vote share as the dependent variable, we also exclude observations where eligible mayors choose not to run. To ensure that voters are responding to a single school quality signal, we exclude observations where separate IDEB results were published for municipal middle schools as well as municipal primary schools. Appendix A6 presents details of how these and a few other data availability constraints limit our sample.

4.1.2 Estimation and Inference

RDDs require specifying the functional form of the regression on both sides of the cutoff and choosing a bandwidth, i.e., the range of the forcing variable beyond which observations are excluded from the analysis. We follow the common practice of using local linear regression, and apply it to the following estimating equation:

$$Y_{mj} = \alpha + \beta_1 T_{mj} + \beta_2 \tilde{D}_{mj} + \beta_3 T_{mj}\tilde{D}_{mj} + \sum_{g=2}^{3} \gamma_g I[g = j] + \sum_{k=1}^{K} \theta^k X_{mj}^k + \varepsilon_{mj} \qquad (3)$$

Y_{mj} is the electoral outcome of interest (an indicator for whether the incumbent mayor was re-elected, or vote share of the incumbent) for municipality m in period j. T_{mj} is a

treatment indicator: $\mathbb{1}$(IDEB score $\geq$ IDEB target). $\tilde{D}_{mj}$ is the distance to the threshold in the forcing variable after centering it around zero: $\tilde{D}_{mj} = D_{mj} + 0.05$. $\sum_{g=2}^{3} \gamma_g I[g = j]$ is a set of election cycle fixed effects (one of which acts as baseline), included because election cycles act as randomization blocks. $\sum_{k=1}^{K} \theta^k X_{mj}^k$ is an additive set of K controls we include for improving the precision of $\hat{\beta}_1$ (Calonico et al., 2019).[3] ε_{mj} is an error term. If the RDD assumptions hold, β_1 identifies the LATE in Equation 3: $\beta_1 = \hat{\tau}$. We use heteroskedasticity-consistent standard error estimators for inference.

To choose the bandwidth, we use the algorithm proposed by Calonico et al. (2014), which determines an optimal bandwidth that minimizes the mean squared error. We then show the sensitivity of the main results to many alternative bandwidths. We also examine the sensitivity of the results to a 'robust' regression discontinuity model as proposed by Calonico et al. (2014), which uses kernel weights (putting more weight on observations closer to the cutoff) and corrects for potential bias.[4]

4.2 Design 2: Field Experiment in Pernambuco State

Observational research designs such as RDDs are subject to concerns about statistical modeling assumptions. Furthermore, our RDD analysis of the effect of meeting IDEB targets relies on aggregate data, limiting our ability to test mechanisms about how voters process information generated by standardized tests. To circumvent these limitations of the RDD, we rely on a field experiment implemented in the state of Pernambuco in

---

[3]Controls include the vote share of the mayor in the previous election; indicators for whether the mayor belongs to major parties PT, PSDB or PMDB; an indicator for whether the party of the mayor runs; and the municipality's logged population, percent of inhabitants who are poor, and share of public employees who are tenured.

[4]We do not apply kernel weighting in our baseline specification, and 'localize' the regression function using the bandwidth alone, as recommended by Lee and Card (2008, 319).

partnership with the State Accounts Court (*Tribunal de Contas do Estado de Pernambuco* or TCE-PE), the primary accountability institution in the state of Pernambuco. This experiment, described more fully in (omitted self-citation), provided information on municipal performance in the ANA to individuals via flyers and an oral message prior to the 2016 municipal elections. We opted to base our educational performance indicator on the ANA, rather than the better known IDEB, because the results of the latter are released only every two years, and the 2015 results—necessary to measure change during the mayor's term—were not available until shortly after our study went to the field.

4.2.1 Treatment

In contrast to the IDEB, there is no pre-existing, readily interpretable summary measure of ANA performance, so we created one for our experiment. The federal government releases the ANA results for each municipality by reporting the proportion of students that are classified into four categories of increasing performance for both the reading and math portions of the exam. To compute an overall score, we calculated the mean level of performance for reading and math combined. To capture an improvement or decline in test scores potentially attributable to the mayor, we then measured the change in this average score between 2012 and 2014. As demonstrated in Appendix B1, there is substantial variation in the degree to which municipalities change over time on exam performance. To communicate the ANA performance results to voters, we ranked all 185 municipalities in the state according to this change score. In each municipality, we report the overall ranking as well as the percentage of municipalities that scored better or worse.

Information was delivered to voters in the form of a flyer handed out by enumerators during the baseline wave of a panel survey; an example of the flyer is in Appendix B2. Enumerators also summarized the information orally to maximize information retention and facilitate comprehension among illiterate voters.[5] The flyer design was refined

---

[5]To satisfy legal restrictions on the distribution of campaign advertisements, flyers

based on feedback from two rounds of focus groups conducted with voters from three municipalities as well as review by our government partner, the TCE-PE. The front of the flyer bore the logos of the TCE-PE and its affiliated academic institution, the Public Accounts School, and it briefly explained the court's auditing responsibilities. The reverse side conveyed municipality-specific details, including a visual illustration of the ranking with comparative metrics.

4.2.2 Data

The experimental sample consisted of 3,200 adult registered voters in 47 municipalities in the state of Pernambuco where the incumbent mayor was running for reelection in 2016. The sample was stratified by performance on our ANA metric, such that equal numbers of respondents lived in municipalities above and below the statewide median. Respondents were randomly assigned with equal probability to a treatment group that received information about ANA performance, a pure control group that received no information, and a second treatment group that received information about the results of an audit of municipal finances by the TCE-PE, which is analyzed elsewhere (omitted self-citation). Assignment was block-randomized at the census tract level.

Our outcome variable, *Vote*, was measured during a second wave of the survey that was fielded 2–4 weeks after the election and reinterviewed 2,577 respondents. Vote takes on a value of 1 if the respondent reported voting for the incumbent mayor, and 0 otherwise (including abstention or a blank or null vote). Nonresponse was not an issue; only one person refused to answer. To reduce social desirability bias and demand effects, we

---

were carefully designed to not meet the definition of campaign advertising according to Brazilian law; they said nothing about elections, voting, or specific candidates. Furthermore, they were reviewed and approved not only by the Ethics in Research Committee (Comitê de Ética em Pesquisa, CEP) of a Brazilian institution, but also by lawyers at the TCE-PE. This issue is discussed further in Appendix B3

used municipality-specific printed ballots, which respondents were asked to deposit in an envelope carried by the enumerator. Brazil uses electronic voting, so it was impossible to mimic the design of an actual ballot, but our paper ballots included all of the information displayed on the electronic voting confirmation screen: name, candidate number, party, and a black and white photo. We also included a space to indicate a blank or null vote, as is possible with electronic voting. We provide an example of the ballot in Appendix B4.

4.2.3 Estimation and Inference

In contrast to the binary IDEB signal used in the RDD, the information presented in the field experiment is continuous in nature. We expect that the effect of providing information about school performance on voting behavior will vary with positivity or negativity of the performance signal. Hence, our main specification involves interacting a binary treatment indicator with the municipality's rank on our ANA performance metric as conveyed in the treatment information. Specifically, we estimate treatment effects using the following equation:

$$Y_{im} = \beta_0 + \beta_1 \mathbf{T}_{im} + \beta_2 \mathbf{R}_m \cdot \mathbf{T}_{im} + \sum_{j=1}^{k} \left(\mu_k X_{im}^K + \gamma_k X_{im}^K \cdot \mathbf{T}_{im} \right) + \epsilon_{im} \tag{4}$$

Y_{im} is the outcome variable for individual i in municipality m, $\mathbf{T_{im}}$ is the treatment indicator, $\mathbf{R}_m$ is the municipal ranking, X_{im}^K is the kth pre-treatment covariate, and ϵ_{im} is the disturbance term. X_{im}^K and $\mathbf{R}_m$ are demeaned using the sample average. Because we demean the covariates and include their interaction with treatment, β_1 is a consistent estimator for the average treatment effect.[6] β_2 is the coefficient on the interaction between the treatment and municipal performance. For the standard error of our estimates, we employ the "HC2" heteroskedastic consistent estimator.

[6]The main effect of the ranking variable is omitted because it is perfectly collinear with block dummies.

To increase the precision of our main estimates, we control for a vector of pre-treatment covariates, in addition to block fixed effects. We employ a *pre-specified* data-adaptive procedure that selects a small number of covariates from all available pre-treatment covariates based on how well they predict the outcome. By using a procedure that optimizes for out of sample predictive performance, we sought to maximize the efficiency of our estimates. Specifically, we follow Bloniarz et al. (2016) and use the "least absolute shrinkage and selection operator" (Lasso) to select a parsimonious set of relevant covariates to include in our estimating equation for each specification. We estimate separate Lasso models in each treatment and control group. We then employ 10-fold cross-validation on the combination of the Lasso and OLS to select optimal tuning parameters for out of sample prediction. Finally, the non-zero coefficients in the Lasso model using the optimal tuning parameter are used in our main estimating equations. Results without covariate adjustment are presented in the Appendix.

As an alternative to a specification with a linear interaction and covariates, we also split the sample into performance terciles and estimate the treatment effect separately in each bin without any covariates (aside from block dummies). This binned approach helps diagnose potential violations of the linearity assumption, as suggested by Hainmueller et al. (2019), and shows that our overall conclusions do not depend on covariate adjustment.

Estimating treatment effects conditional on our ANA performance ranking departs from our pre-specified approach. In concert with the broader initiative of which the field experiment was a part, we hypothesized that the effect of school performance information would vary based on whether it was "good news" or "bad news" (measured dichotomously) in comparison to a respondent's priors. In retrospect, this approach does not work well for our measure. First, the granularity of the underlying ranking can lead to counterintuitive binary classifications—someone who guesses their municipality is ranked last (185th place) in the state and is told that it ranks next-to-last would be

17

scored as receiving good news, when in reality their highly negative prior has essentially been confirmed. Second, people are unlikely to have well-informed priors about a ranking that was constructed for this project and had never been communicated in the media. The correlation of true ANA rank and priors on this measure is 0.17, and 20% of the sample gave a "don't know" response, suggesting that priors are noisy and conditioning on them would simply generate inefficiency. That said, using the pre-specified approach (reported in Appendix B10), the overall effect of receiving "good news" relative to priors is similar to what we estimate below for respondents from the best-ranked municipalities.

In addition to estimating treatment effects conditional on ANA performance ranking, we present separate results for those respondents who have children enrolled in municipal schools, for whom we expect the treatment information to be more salient. This particular hypothesis goes beyond the pre-analysis plan, though it is consistent with our general pre-registered expectation that "the effect of information provision on voting behavior will depend on the salience of the corresponding policy area for individual welfare" (omitted self-citation).

5   Results

5.1   RDD results

Table 1 shows the main results of the regression discontinuity design. Contrary to expectations, the incumbent's chances of re-election are *lower* in municipalities that met their IDEB target than in those that missed it. In our preferred specification (using local linear regression with controls), around the threshold meeting the IDEB target decreases the probability of incumbent re-election by 8.5 percentage points ($p < 0.05$) or over 17% of a standard deviation. This result is visualized in Figure 1. The "robust" specification of Calonico et al. (2014) returns a similar estimate with slightly larger standard errors ($p = 0.052$). Results examining the effect on vote-share (included in Appendix A7) are

similar, if noisier due to a lower number of observations.

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.081*	-0.085**	-0.098*	-0.109*
	0.045	0.043	0.054	0.056
Election cycle fixed effects	✓	✓	✓	✓
Controls		✓		✓
Bandwidth	0.393	0.383	0.393	0.383
N	1795	1755	1795	1755

*p<0.1; **p<0.05; ***p<0.01.

Table 1 – Effect of Meeting the IDEB Target on Re-election of the Mayor. The bandwidth is the one determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4.)
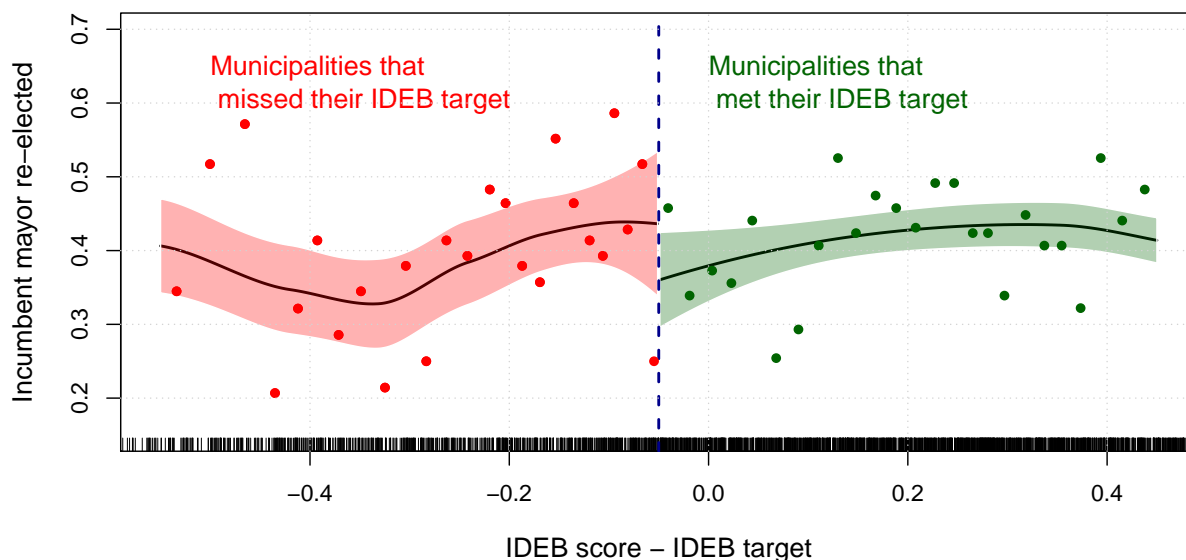
Figure 1 – Effect of Meeting the IDEB Target on Re-election of the Mayor. Colored dots represent local averages for 50 equally-sized bins. Lines are loess regression lines estimated at both sides of the threshold with no controls. Shaded regions are their 95% confidence intervals.

One concern with RDD results is that they may be dependent on the choice of a particular bandwidth. As shown in Appendix A8, results in model 2 have limited robustness to the choice of alternative bandwidths. While some of the estimates have 95% confidence intervals that cross 0, point estimates remain large and relatively stable across a

wide range of alternatives to the bandwidth specified by the Calonico et al. (2014) algorithm. As additional robustness checks, we ran a number of placebo tests by moving the RD threshold away from the point where IDEB targets are met. Results, which are shown in Appendix A9, show that all placebo tests return a statistically insignificant result at the conventional 95% level.

Summing up, the RD design shows that meeting the IDEB target has a negative effect on the electoral performance of the mayor—voters appear to punish, rather than reward, the delivery of quality schools. While the significance of the results is not always robust to the choice of bandwidth or specification, the magnitude and sign of the estimated treatment effect is stable across specifications. Moreover, the RD design passes placebo tests and there is no evidence that assumptions are violated, lending support to the interpretation of these findings as causal effects.

5.2   Experiment Results

Treatment effect estimates from the field experiment in Pernambuco are presented in Figure 2. The individual-level experimental evidence aligns with the RDD findings: when informed about their municipality's ranking on the ANA, voters punish good performance. The black line represents the estimate of treatment effect heterogeneity using a regression model with a linear interaction, while the points and vertical lines show the treatment effect estimated separately in each tercile of ANA rank. Contrary to expectations, we find that voters in higher performing municipalities (rank closer to 1) punish incumbents more when receiving the information than those living in municipalities with lower performance (rank closer to 185). While the interaction is imprecisely estimated, the pattern is consistent using both the linear interaction and the tercile approach. Point estimates and standard errors both for the average treatment effect and the linear interaction can be found in Table 2 (column 1).

If voters in poor performing municipalities had lower expectations of their mayor

20

than voters in higher performing municipalities, the positive interaction we find might be driven by Bayesian updating, as described in Arias et al. (2018). As we show in Appendix B8, however, the gap between ANA ranking and priors on this measure is uncorrelated with treatment effect size, making this alternative explanation unlikely to hold.
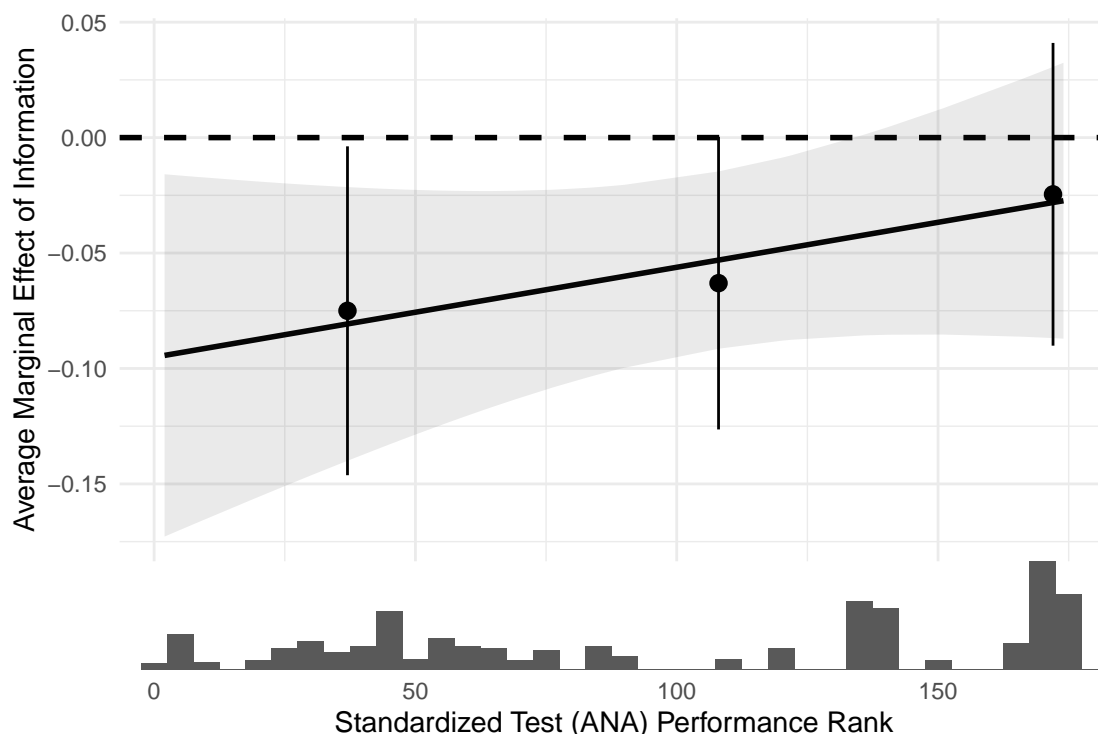

Figure 2 – Effect of Treatment by Educational Performance. Black line is estimated marginal effect of treatment estimated using a linear interaction. Points are effects estimated separately in bins defined by the terciles of ANA Rank. 95% Confidence Intervals are shown. Histogram shows marginal distribution of ANA Rank.

While estimated treatment effects in the full sample run counter to our expectations, it is possible that parents of children enrolled in municipal schools, for whom the treatment information should be particularly salient, react in a different manner. Figure 3 displays the estimated effect of the information among parents with children enrolled in municipal schools versus the rest of the sample. Disaggregating the data in this fashion reveals considerable heterogeneity: those with children in local schools punish poor performers (left figure), while the rest of the sample punishes good performers (right figure). Hence, among this subgroup, our theoretical expectations about the effect of information
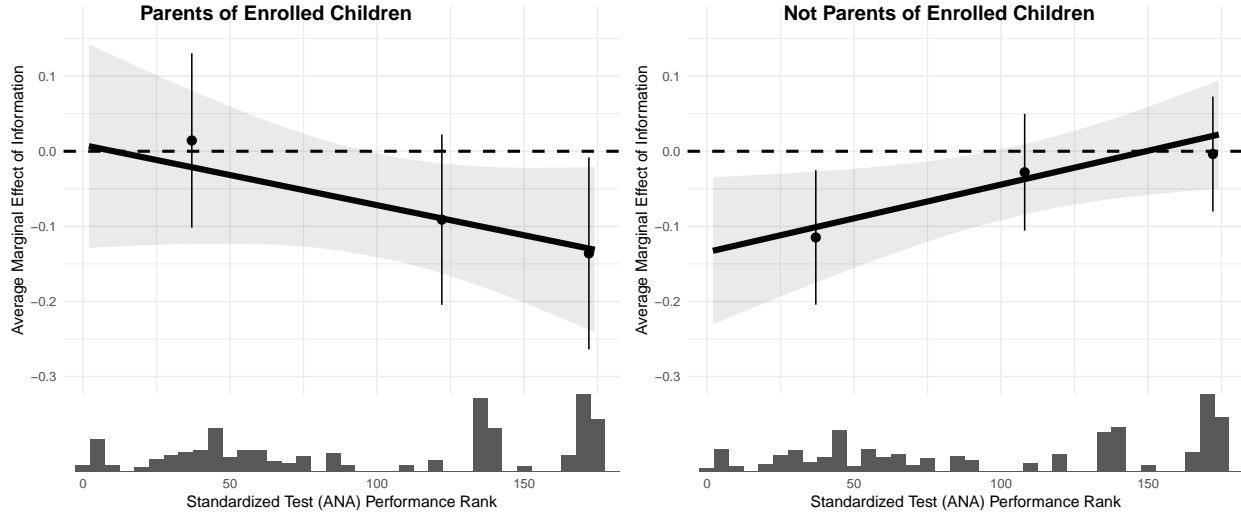
Figure 3 – Effect of Treatment Among Parents and Non-Parents. Black line is estimated marginal effect of treatment estimated using a linear interaction. Points are effects estimated separately in bins defined by the terciles of ANA Rank. 95% Confidence Intervals are shown. Histogram shows marginal distribution of ANA Rank.

on voting behavior are upheld. For respondents without a child in municipal schools, the slope of the linear interaction is significant and positive (column 3 in Table 2). We obtain an insignificant estimate for the interaction term in the parents subgroup, likely due to its smaller size (column 2 in Table 2). However, the difference in slopes between the two groups is statistically significant at the 5% level (column 4 of Table 2). Hence, there is clear evidence that information about school quality has a different effect on voting behavior among parents of children enrolled in municipal schools.

This striking contrast between parents and non-parents does not appear to be driven by observable characteristics or perceptions that are correlated with having children in municipal schools. Parents are somewhat less educated, younger, and poorer than non-parents, but including these variables as additional interactions does not change the relationships observed in Figure 3 (see Appendix B7).

In sum, results from the field experiment are consistent with the RDD: when informed about standardized test scores in their municipality, voters punish good performance by voting against the incumbent mayor. Yet in this individual-level analysis, we are able

22

	All	Parents	Not Parents	All
Treatment	-0.0557^{**}	-0.0707^*	-0.0421	-0.0472^*
	(0.0196)	(0.0350)	(0.0239)	(0.0200)
Treatment x Rank	0.0004	-0.0008	0.0009^*	0.0008
	(0.0003)	(0.0006)	(0.0004)	(0.0004)
Treatment x Rank x Parents				-0.0015^*
				(0.0007)
Num. obs.	1709	525	1184	1709

$^{***}p < 0.001, ^{**}p < 0.01, ^*p < 0.05$

Table 2 – Effect of Information Treatment. Covariates omitted. Covariates demeaned so treatment coefficient is estimated average treatment effect. HC2 heteroskedasticity consistent standard errors in parentheses.

to document a revealing form of heterogeneity: parents of children enrolled in municipal schools react to information about school quality as expected, punishing poor-performing mayors.

6   Testing the Mechanism: Online Survey Experiment

While it is reassuring that parents of children in municipal schools react as expected to information about school quality, the results of the field experiment do not explain why other respondents react in the opposite fashion, punishing mayors in places where standardized test scores improved over time. One possibility is suggested by Bursztyn (2016), who shows that poor voters in Brazil react negatively to information about increased educational spending because they prefer that money be spent on conditional cash transfers. If voters believe that positive (negative) school performance implies that more (less) money is being spent on education, they might punish or reward the mayor based on the implications for those areas of spending that they value the most.

6.1 Research Design

To test whether voters draw connections between school quality and educational spending, we draw on an online survey experiment conducted in September 2018.[7] Our treatment informed respondents about whether their self-reported municipality had met its IDEB target when data were released in September 2016. Immediately prior to the experiment, all respondents were presented with basic information about IDEB and what it means to meet or miss the target, and they were asked if they had heard of it before. Those randomized into the control group received no further information about IDEB. For those in the treatment group, the text of the next question was:

> In the year 2016, at the end of the term of Mayor [Name], [Municipality] [achieved/did not achieve] its IDEB target. Did you hear about that result?

Information in brackets corresponded to actual data from the respondent's self-reported municipality. The text was accompanied by two images designed to reinforce the message: a photograph of the mayor and a red cross or a green checkmark, depending on the IDEB result.[8]

We examine the effect of this treatment on assessments of municipal education spending. The next battery of questions asked respondents whether they totally agreed, partially agreed, partially disagreed, or totally disagreed with a series of statements, one of which was that the mayor "invested a lot of money in education." In fact, as shown in Appendix D, there is no empirical relationship between change in educational spending

---

[7]Respondents were recruited via Facebook ads, a common method for online surveys in comparative politics (Boas et al., 2018). See Appendix C1 for details.

[8]In those cases where a new mayor was elected in October 2016 and took office in January 2017, the treatment information and outcome measure pertain to the previous mayor. There should have been little ambiguity given that we used mayors' names and photographs.

and change in IDEB scores or the likelihood of meeting the target (see also Monteiro, 2015; Menezes Filho et al., 2009). However, we expect that voters will nonetheless connect the two, given the common tendency to assume that policy outputs are related to policy inputs. Hence, informing about meeting (missing) the IDEB target should make voters more (less) likely to agree that the mayor invested a lot of money in education.[9]

6.2 Estimation and Inference

To analyze the impact of the information treatment on assessments of municipal education spending, we use the following equation, estimated separately for respondents from municipalities where the IDEB target was and was not met:[10]

$$Y_i = \alpha + \beta T_i + \sum_{k=1}^{K} \left(\theta_k X_i^k + \gamma_k X_i^k T_i \right) + \varepsilon_i \tag{5}$$

Y_i is the outcome variable (a 4-point Likert scale for each statement about the mayor, with higher numbers indicating agreement) for respondent i and T_i is an indicator for whether they get treated. We include K covariates (sex, gender, race, education level, and region fixed effects) to increase precision. To ensure that $\hat{\beta}$ consistently estimates the treatment

---

[9]Due to an oversight, we failed to include this specific hypothesis in the registered pre-analysis plan for the online survey, though we did list the question as an outcome variable. All pre-specified results are presented in Appendix C4.

[10]In the pre-analysis plan, we had stated we would analyze all data together, with separate treatment indicators for positive IDEB results and treatment status. We prefer the split-sample approach, since it ensures that treated respondents are compared to control-group respondents from the same set of municipalities that either met or did not meet the IDEB target. Nonetheless, results using the pre-specified model, which are reported in Appendix C4, are similar.

effect, we demean all covariates X_i^k and interact them with the treatment indicator. We use the HC2 estimator of the standard errors to account for heteroskedasticity.

6.3 Results

The results of the survey experiment present clear evidence that voters associate signals of school quality with spending on education. Being told that the municipality met its IDEB target boosts agreement with the statement that the mayor "invested a lot of money in education" by 0.11 points on the 4-point scale; being told that the municipality missed its IDEB target decreases agreement with that statement by 0.20 points. Both effects are statistically significant at the 0.05 level or better.

	Positive information	Negative information
	(1)	(2)
Treatment	0.110*	−0.204**
	(0.045)	(0.072)
N	2,173	846
R^2	0.017	0.051

*p<0.05; **p<0.01; ***p<0.001

Table 3 – Effect of receiving positive or negative information about the quality of schools on respondents' agreement with the statement the mayor "invested a lot of money in education", where 1 = disagree completely, and 4 = agree completely. HC2 heteroskedasticity consistent standard errors in parentheses. Controls omitted from the table.

While we did not test whether positive school quality signals lead voters to believe that spending *decreased* in other areas, it is likely that they make this inference. Municipal budgets often run short in Brazil, especially during the recession of 2015–2016, which saw an annual GDP growth rate of –3.5%. In a November 2018 survey of Brazilian municipalities, 83% responded that they had sought to deal with the crisis by reducing operating expenses in the municipal bureaucracy, and 54% had resorted to reducing personnel expenditures (Confederação Nacional dos Municípios, 2018). These cutbacks are likely to have impacted people and their friends and family directly. One in 22 Brazilians ages

18–64 is employed by a municipal government, and the average municipality devotes more than 50% of its budget to personnel salaries. In a context of austerity, one would expect budget constraints to be at the top of people's minds and expanded government spending in one area to imply cutbacks in others.

Our findings regarding the causal mechanism are suggestive rather than conclusive, but they imply that most Brazilian voters punish positive educational performance because they perceive trade-offs with spending in other, higher priority areas. Our results suggest that trade-off thinking can be triggered not only by providing information about policy inputs, as in the experiment by Bursztyn (2016), but also by informing voters about policy outputs.[11] These findings also echo recent research on European welfare states showing that while education is a valence issue, citizens' preferences for education spending (both in absolute terms and relative to other policy areas) vary systematically across socioeconomic groups in line with trade-off thinking (Busemeyer et al., 2018; Busemeyer and Garritzmann, 2017).

7   Discussion and Conclusion

Governmental and non-governmental agencies around the world are increasingly turning to transparency and information campaigns in attempts to foster electoral accountability. By providing information on government performance, these initiatives hope to enable and empower citizens to reward elected officials who deliver high-quality public services and to punish those who do not perform well. But does this logic actually hold, in practice?

Cutting-edge research on electoral accountability suggests that the link between in-

---

[11]Our argument also differs from that of Bursztyn (2016) in that he focuses on and finds evidence of trade-offs only among the poor. In contrast, we do not find evidence that our effect is driven by poor voters (see Appendix A11).

formation and accountability is weak (Dunning et al., 2019,?), and it critically depends on institutional, socioeconomic and behavioral features, including prior beliefs (Arias et al., 2018), expectations (Gottlieb, 2016), socioeconomic endowments (Holbein, 2016), coordination (Adida et al., 2017), media markets (Larreguy et al., 2015), and ethnicity (Adida et al., 2017). Empirically, most of these studies use field experiments, which maximize inferential rigor at the expense of external validity and make it hard to measure potential general equilibrium effects.

To address these theoretical and empirical issues, we approach the link between information and electoral accountability through a two-fold study of information and accountability in the Brazilian education sector. First, we use a regression discontinuity design, which effectively compares municipalities that barely met a simple and highly visible indicator of municipal school quality to municipalities that barely missed it. We thus measure the relationship between school quality signals and electoral outcomes at the macro level, leveraging data for all Brazilian municipalities. Second, we use a field experiment in the state of Pernambuco, which examines the effect of providing individuals with information about the relative performance of the schools in their municipality. Together, our research designs allow us to strengthen our inferences while addressing issues of both internal and external validity.

We find a consistent result across the two designs and measures of school quality: positive information decreases the incumbent's electoral performance. In the field experiment, the effect is reversed for parents of children enrolled in municipal schools, for whom school quality should be most salient. To test the psychological mechanism behind these results, we conducted an online survey experiment. We find that voters tend to infer a relationship between school quality and municipal spending on education, even though no such relationship exists in the real world. Given this belief, they are also likely to assume that positive educational performance implies reduced spending in other areas. Hence, those who do not directly benefit from school quality may choose to punish

rather than reward mayors for this form of service provision.

Our findings add to an emerging body of literature on "accountability backlash"—the notion that interventions designed to increase electoral accountability by equipping voters to sanction poor performing politicians and reward good ones can can end up having the opposite effect. Prior studies have identified a variety of mechanisms that can underpin negative electoral responses to public service delivery, including Bayesian learning (Arias et al., 2018), voter demobilization (Chong et al., 2015), perceived tradeoffs between programmatic performance and particularistic transfers (Adida et al., 2017), increased expectations, and exposure to negative side effects of service delivery such as corruption (De Kadt and Lieberman, 2017). We add to this literature by arguing that voters may also perceive trade-offs between different programmatic policy areas, and that policy outputs can induce trade-off thinking by prompting inferences about their associated inputs.

The findings have important policy implications. First, policymakers should give serious consideration to the potential heterogeneous treatment effects of information campaigns and transparency systems. Not all voters hold the same preferences, and responses to information may differ systematically across the the electorate. In some cases, heterogeneity may mean that information prompts accountability voting only for some small subgroup of the population or under an unusual set of circumstances, and it does not work on a broad enough scale to actually affect politicians' electoral prospects or induce them to perform better in the future. Second, policymakers should be aware that performance-based accountability systems may backfire, leading to effects in the opposite direction of those that were intended. While accountability backlash may affect only a portion of the electorate—as we find in the case of citizens who do not have children enrolled in municipal schools—that portion may be sufficiently large, or backlash effects may be of sufficient magnitude, that aggregate effects of an intervention go in an undesired direction. Third, policymakers and researchers alike should pay more attention to the potential trade-offs that voters may perceive between different policy outputs, as they

can seriously compromise the success of information and transparency initiatives.

References

Adida, C., J. Gottlieb, E. Kramon, and G. McClendon (2017). Breaking the clientelistic voting equilibrium: The joint importance of salience and coordination. *Working Paper*.

Adida, C., J. Gottlieb, E. Kramon, G. McClendon, et al. (2017). Reducing or reinforcing in-group preferences? an experiment on information and ethnic voting. *Quarterly Journal of Political Science 12*(4), 437–77.

Arias, E., H. Larreguy, J. Marshall, and P. Querubin (2018). Priors Rule: When do Malfeasance Revelations Help or Hurt Incumbent Parties. *National Bureau of Economic Research*.

Banerjee, A., S. Kumar, R. Pande, and F. Su (2011). Do Informed Voters Make Better Choices? Experimental Evidence from Urban India. *Working Paper*.

Barber, M., N. C. Rodriguez, and E. Artis (2015). *Deliverology in Practice: How Education Leaders are Improving Student Outcomes*. Corwin.

Blattman, C., M. Emeriau, and N. Fiala (2018). Do Anti-Poverty Programs Sway Voters? Experimental Evidence from Uganda. *Review of Economics and Statistics 100*(5), 891–905.

Bloniarz, A., H. Liu, C.-H. Zhang, J. S. Sekhon, and B. Yu (2016). Lasso Adjustments of Treatment Effect Estimates in randomized Experiments. *Proceedings of the National Academy of Sciences 113*(27), 7383–7390.

Boas, T. C., D. P. Christenson, and D. M. Glick (2018). Recruiting Large Online Samples in the United States and India: Facebook, Mechanical Turk and Qualtrics. *Political Science Research and Methods*, 1–19.

Boas, T. C. and F. D. Hidalgo (2019). Electoral Incentives to Combat Mosquito-Borne Illnesses: Experimental Evidence from Brazil. *World Development 113*, 89–99.

Boas, T. C., F. D. Hidalgo, and M. A. Melo (2019). Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil. *American Journal of Political Science 63*(2), 385–400.

Bruns, B., D. Evans, and J. Luque (2012). *Achieving World-Class Education in Brazil: The Next Agenda*. World Bank Publications.

Buntaine, M. T., R. Jablonski, D. L. Nielson, and P. M. Pickering (2018). SMS Texts on Corruption Help Ugandan Voters Hold Elected Councillors Accountable at the Polls. *Proceedings of the National Academy of Sciences 115*(26), 6668–6673.

Bursztyn, L. (2016). Poverty and the Political Economy of Public Education Spending: Evidence from Brazil. *Journal of the European Economic Association 14*(5), 1101–1128.

Busemeyer, M. R. and J. L. Garritzmann (2017). Public opinion on policy and budgetary trade-offs in european welfare states: evidence from a new comparative survey. *Journal of European Public Policy 24*(6), 871–889.

Busemeyer, M. R., P. Lergetporer, and L. Woessmann (2018). Public opinion and the political economy of educational reforms: A survey. *European Journal of Political Economy 53*, 161–185.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). Regression Discontinuity Designs Using Covariates. *Review of Economics and Statistics 3*(101), 1–10.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82*(6), 2295–2326.

Chong, A., A. de la O, D. Karlan, and L. Wantchekon (2015). Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification. *Journal of Politics 77*(1), 55–71.

Confederação Nacional dos Municípios (2018). O pagamento do 13° Salário pelos Municípios brasileiros em 2018. Estudo Técnico, December.

Cunow, S. and S. Desposato (2015). Local review: Confronting the brazilian black box. In S. Desposato (Ed.), *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, pp. 128–138. New York: Routledge.

De Figueiredo, M. F. P., F. D. Hidalgo, and Y. Kasahara (2011). When do Voters Punish Corrupt Politicians? Experimental Evidence from Brazil. Working paper, http://bit.ly/2ByZMqK.

De Kadt, D. and E. S. Lieberman (2017). Nuanced Accountability: Voter Responses to Service Delivery in Southern Africa. *British Journal of Political Science*, 1–31.

Desposato, S. (2015). Introduction. In S. Desposato (Ed.), *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, pp. 1–22. New York: Routledge.

Dunning, T., G. Grossman, M. Humphreys, S. Hyde, C. McIntosh, and G. Nellis (2019). *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. New York: Cambridge University Press.

Dunning, T., G. Grossman, M. Humphreys, S. Hyde, C. McIntosh, G. Nellis, C. L. Adida, E. Arias, C. Bicalho, T. C. Boas, M. T. Buntaine, S. Chauchard, A. Chowdhury, J. Gottlieb, F. D. Hidalgo, M. Holmlund, R. Jablonski, E. Kramon, H. Larreguy, M. Lierl, J. Marshall, G. McClendon, M. A. Melo, D. L. Nielson, P. M. Pickering, M. R. Platas, P. Querubin, P. Raffler, and N. Sircar (2019). Voter Information Campaigns and Political Accountability: Cumulative Findings from a Pre-Registered Meta-Analysis of Coordinated Trials. *Science Advances 5*(7), eaaw2612.

Fernandes, R. (2007). Índice de Desenvolvimento da Educação Básica (IDEB): Metas Intermediárias para a sua Trajetória no Brasil, Estados, Municípios e Escolas. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília*. Available at http://download.inep.gov.br/educacao_basica/portal_ideb/o_que_sao_as_metas/Artigo_projecoes.pdf.

Ferraz, C. and F. Finan (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics 123*(2), 703–745.

Gaventa, J. and R. McGee (2013). The Impact of Transparency and Accountability Initiatives. *Development Policy Review 31*(s1), s3–s28.

Gottlieb, J. (2016). Greater Expectations: A Field Experiment to Improve Accountability in Mali. *American Journal of Political Science 60*(1), 143–157.

Hainmueller, J., J. Mummolo, and Y. Xu (2019). How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. *Political Analysis 27*(2), 163–192.

Holbein, J. (2016). Left Behind? Citizen Responsiveness to Government Performance Information. *American Political Science Review 110*(2), 353–368.

Humphreys, M. and J. Weinstein (2012). Policing Politicians: Citizen Empowerment and Political Accountability in Uganda: Preliminary Analysis. Working paper, http://bit.ly/2kmrljT.

Imbens, G. W. and T. Lemieux (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of econometrics 142*(2), 615–635.

Larreguy, H., J. Marshall, and J. M. Snyder Jr (2015). Publicizing Malfeasance: When Media Facilitates Electoral Accountability in Mexico. *Working paper*.

Lee, D. S. and D. Card (2008). Regression Discontinuity Inference with Specification Error. *Journal of Econometrics 142*(2), 655–674.

Lee, D. S. and T. Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature 48*, 281–355.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics 142*(2), 698–714.

Menezes Filho, N. A., L. Amaral, et al. (2009). A relação entre gastos educacionais e desempenho escolar. *XXXVI Encontro Nacional de Economia. Salvador (Bahia) 9*.

Monteiro, J. (2015). Gasto público em educação e desempenho escolar. *Revista Brasileira de Economia 69*(4), 467–488.

Pande, R. (2011). Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies. *Annual Review of Economics 3*(1), 215–237.

Weitz-Shapiro, R. and M. S. Winters (2017). Can Citizens Discern? Information Credibility, Political Sophistication, and the Punishment of Corruption in Brazil. *The Journal of Politics 79*(1), 60–74.

Appendices

A   Regression discontinuity design

A.1   Quantitative measures of media attention to IDEB

In order to get a quantitative measure of media attention, we use the search tool of written media content aggregator Factiva, looking for news in Portuguese from within Brazil mentioning IDEB, between the date when the data was published and the day of the elections. For 2016, this search returns 254 pieces of news, more than half of which also included the word "target" or "targets". This can be compared to a baseline of over 52,000 pieces in the Factiva database for that same period, i.e. about 0.49% of the news pieces in that period mention IDEB. For the dates between the publication of IDEB results and municipal elections in 2012, searching for news mentioning IDEB returns 494 pieces of news, 37% of which include the word "target" or "targets". For 2008, the search returns 239 pieces.

This supports the idea that media coverage emphasizes not just levels of achievement but also and particularly how they relate to the targets. While Factiva covers a large number of sources, it does not cover the universe of Brazilian media, and it is thus hard to judge how much coverage 254 search hits represent. To get a benchmark, we look for news pieces about Bolsa Família, the federal government's highly visible and often discussed conditional cash transfer program. For the same period from the publication of IDEB results up to election day in 2016, Factiva returns 215 pieces of news mentioning Bolsa Família. Overall, it seems Brazilian written media pay significant attention to IDEB. We have no evidence on coverage by radio and television, but given the prominence of IDEB in newspapers, it is reasonable to expect that TV and radio stations cover it extensively as well.

A.2 Quantitative measures of citizen attention to IDEB

We use Google Trends data on searches made from within Brazil as a proxy for citizen demand for information about IDEB. Figure A.1 below compares the relative frequency of searches for "IDEB" and a number of highly salient policy issues.
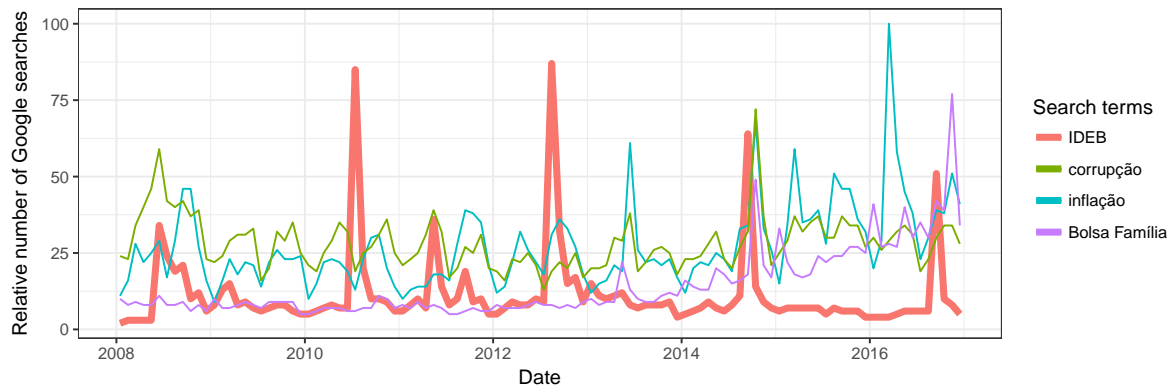
Figure A.1 – Relative frequency of Google searches in Brazil for the terms "IDEB", "corruption", "inflation", and "Bolsa Familia" from 2008 to 2016, by month. Data are from Google Trends.

A.3 Information release and election schedule
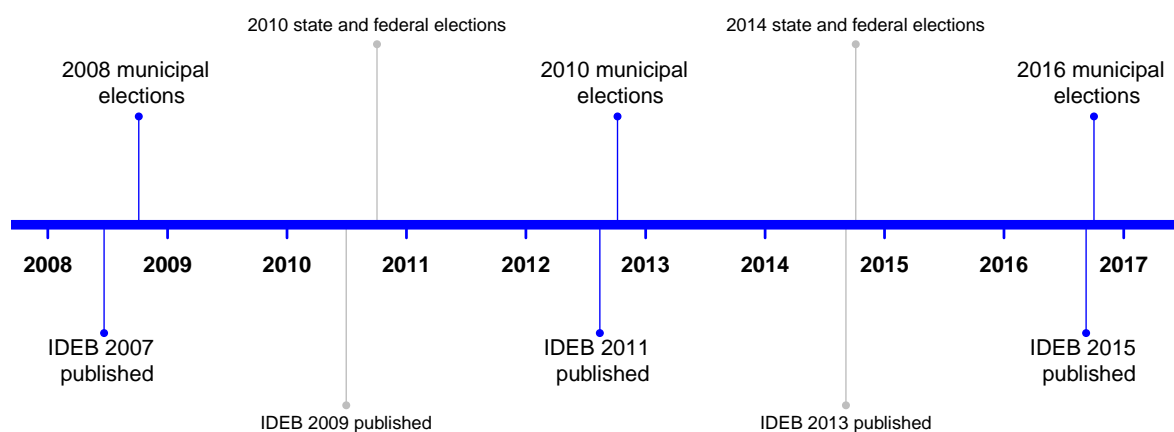
Figure A.2 – Timeline of the release of IDEB results and elections. See Table A1 below for dates of each event.

ANRESC implemented	IDEB results published	First round of elections held	Time from results to elections
November 5-20, 2007	June 21, 2008	October 5, 2008	<18 weeks
October 19-30, 2009	July 1, 2010	October 3, 2010	<14 weeks
November 7-18, 2011	August 14, 2012	October 7, 2012	<8 weeks
November 11-21, 2013	September 5, 2014	October 5, 2014	<5 weeks
November 3-13, 2015	September 8, 2016	October 2, 2016	< 4 weeks

Table A1 – Key dates in the IDEB and the electoral calendar. Municipal elections held in 2008, 2012 and 2016. State and federal elections held in years 2010 and 2014.

5

A.4 Continuity of the forcing variable
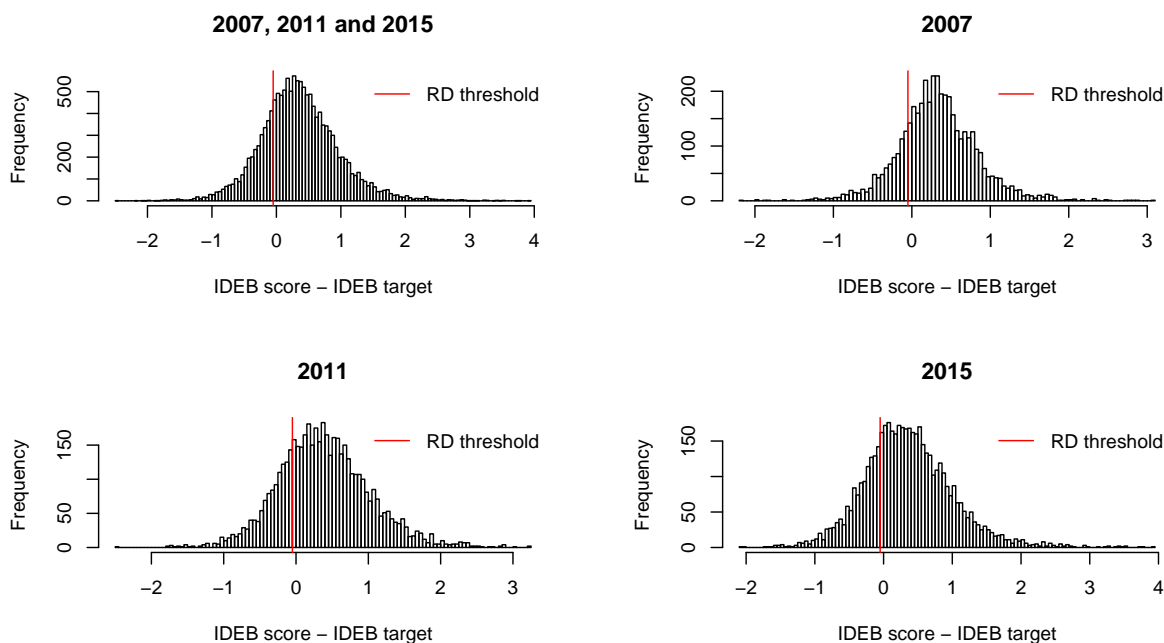
Figure A.3 – Histograms of the forcing variable, by test year.

While "a running variable with a continuous density is neither necessary nor sufficient for identification" (McCrary, 2008, 701), it is important to consider possible ways teachers, directors and politicians could be manipulating the forcing variable. IDEB targets are impossible to manipulate. They were defined a priori following technical criteria and published at the beginning of the period. IDEB scores are themselves composed of two parts: passing rates and learning outcomes. Passing rates are the most obvious lever that school and municipality leaders could manipulate. However, boosting passing rates is likely to lead to a decrease in test scores (since students who would otherwise not pass generally get lower scores): the system is in fact designed to disincentivize this type of manipulation. Last, learning outcomes are under *limited* control of school administrators and teachers. IDEB is precisely targeted at measuring their capacity of "manipulating" this variable, i.e. boosting learning. But boosting learning is difficult, and even units that manage to achieve significant gains in learning may miss their target, particularly if they had been lagging behind. The key fact here is that while teachers, directors and
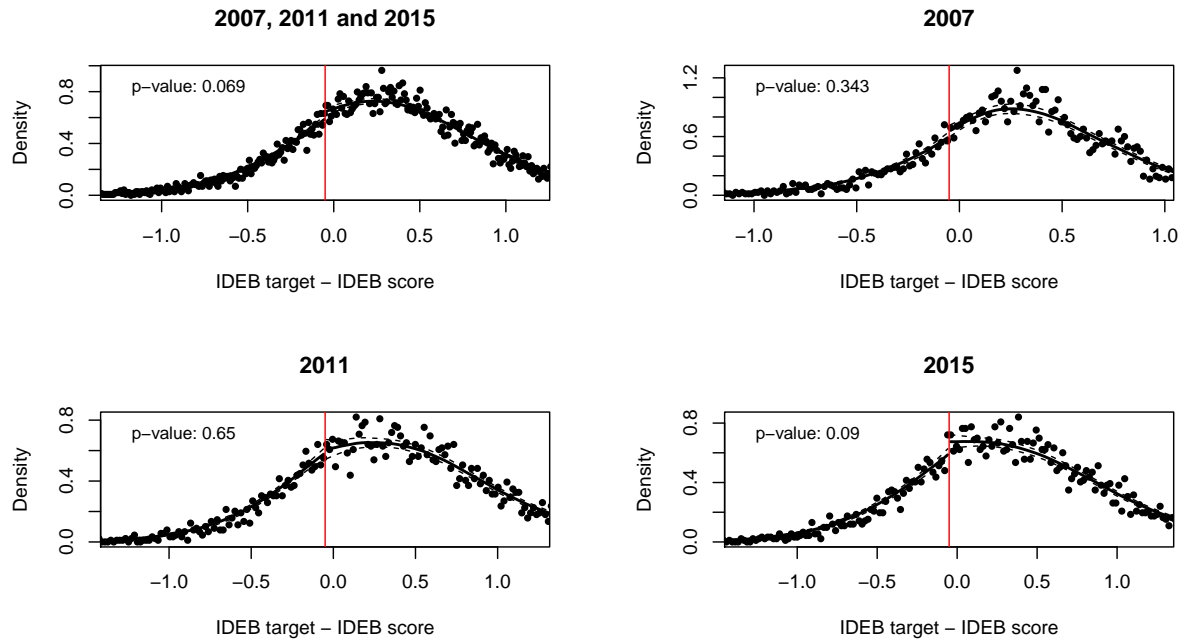
Figure A.4 – Density plots and results of the McCrary (2008) test for the forcing variable (IDEB score - IDEB target), by test year.

politicians may have some influence over the forcing variable, they cannot manipulate it *precisely*, which guarantees that, for municipalities around the threshold, treatment assignment is as-if-random (Lee and Lemieux, 2010).

A.5 Continuity of pre-treatment covariates

	RD estimate of of T-C	Standard error	p-value
Mayor vote share in previous election	0.001	0.007	0.924
Mayor re-elected in previous election	-0.011	0.028	0.688
Incumbent mayor belongs to PT	0.002	0.014	0.873
Incumbent mayor belongs to PSDB	-0.033	0.021	0.109
Incumbent mayor belongs to PMDB	0.022	0.020	0.280
Population	-13025.481	14604.084	0.372
Household monthly per capita income 2010	-1.942	10.051	0.847
Families in the Bolsa Família program	-289.196	420.238	0.491
Share of municipal employees who are civil servants	0.000	0.010	0.974
Radio & TV stations 2012	0.019	0.055	0.734
IDEB score	0.084	0.049	0.084
Share of primary school enrollments in municipal schools	-0.003	0.010	0.795

Table A2 – Continuity of pre-treatment covariates. The difference between municipalities where the IDEB target is met and those where it is missed is calculated using each covariate as the dependent variable in Equation 3, within the bandwidth specified by the Calonico et al. (2014) algorithm. Standard errors are consistent for heteroskedasticity (HC1).

A.6 Data constraints and characterization of the RDD effective sample

With 5,565 municipalities until 2012, and 5,570 thereafter, we have 16,700 potential municipality-election observations. Two reasons constrain the number of observations we can use for the RDD.

- *Forcing variable.* First, our forcing variable is not defined for all municipalities. At every wave, a small proportion of the municipalities do not have a primary education IDEB score published, or lack a target. As a result, the forcing variable is only defined for a total of 14,240 municipality-election observations.

    - IDEB score. 579 municipalities lack a primary education IDEB score for 2007, 429 for 2011, and 582 for 2015. By far the most common reason for a municipality to lack an IDEB score is that the municipality does not have any school with at least 20 students enrolled in the corresponding grade, which INEP requires in order to calculate the ANRESC scores. Other reasons include some special circumstances in which the Ministry allows some municipalities apply for a waiver of the publication of their results, or municipalities where less than half of the students enrolled in the corresponding grade actually sit the exam. These cases are rare: only 42 municipalities in 2011 and 50 in 2015 had their results not published because of the 50% rule; and only 10 in 2011 and 11 in 2015 had their results not published because they were granted a waiver.

    - IDEB target. 1,221 municipalities lack a primary education IDEB target for 2007, 464 for 2011, and 255 for 2015. These are due to the fact that the Ministry needs a baseline measurement of learning outcomes through the ANRESC to calculate targets.

- *Dependent variable.* Second, our dependent variables also impose some restrictions on the sample. The RDD has two main dependent variables: an indicator for whether the mayor was re-elected, and the vote share of the incumbent mayor.

    - Mayor eligibility to run. We exclude municipality-election observations where

9

the mayor has had two consecutive terms in office and is thus ineligible for re-election. This happened in 1,258 municipalities in 2008, 2,081 in 2012, and 1,362 in 2016. As a result, the incumbent was allowed to run in only 11,518 municipality-election observations.

– Missing identifier of the mayor. In 38 municipality-election observations the unique identifier of the mayor (which we use to measure incumbent vote share and re-election) is unfortunately not reported in the electoral data.

– We also exclude observations where the elections were declared as invalid and supplementary elections were held (usually months or even years after the original elections), which happened in 276 municipality-election observations.

– The incumbent's vote share is defined only for municipalities where the incumbent actually runs. Incumbents did not run in 8,529 observations. The indicator for incumbent re-election is however defined also for cases in which the mayor does not run, which we simply code with a 0.

• One signal only. Finally, we also exclude the 8,431 municipality-period observations for which the Ministry of Education publishes information about the performance of municipal middle schools, on top of municipal primary schools.

With all these constraints, our effective sample size is 4,179 municipality-period observations for regressions where incumbent re-election is the dependent variable, and 2,868 where incumbent vote share is the dependent variable.

Our results results are therefore valid for the subset of municipality-election observations where (i) there is a public IDEB score and target, (ii) the incumbent mayor is not barred from running (and, for the regressions of vote share, they actually run); (iii) the regular elections are valid and thus no supplementary elections are held; (iv) only one signal of municipal school quality is published; and (v) the IDEB score performance is sufficiently close to the target. While this set does not represent the whole population of municipalities in Brazil, it is the set for which it is meaningful to think of the causal

10

impact of meeting the IDEB target on the electoral performance of the incumbent.

Table A3 characterizes the RDD effective sample in terms of the number of residents (logged), the percent of them who are poor (as defined by the federal statistics index), local political competitiveness (measured with a Herfindahl index of electoral concentration where numbers closer to 0 denote a more fragmented election and those closer to 1 a less competitive election), and the share of municipal workers who are tenured. Compared to all municipality-year observations, those that enter the RDD belong to less populous municipalities, with slightly lower incidence of poverty, and less electoral fragmentation.

	Dependent variable: Observation is included in:		
	RDD sample (compared to all)	RDD sample & within ± 0.4 bandwith (compared to all)	RDD sample & within ± 0.4 bandwith (compared to those in the RDD sample)
	(1)	(2)	(3)
Population (logged)	−0.046***	−0.012***	0.038***
	(0.003)	(0.002)	(0.008)
Percent residents poor	−0.005***	−0.002***	0.0003
	(0.0002)	(0.0001)	(0.001)
Electoral concentration	0.110***	0.042*	−0.039
	(0.024)	(0.018)	(0.051)
Share workers tenured	0.022	0.045***	0.110**
	(0.015)	(0.011)	(0.035)
Constant	0.744***	0.233***	0.063
	(0.038)	(0.028)	(0.092)
Election fixed effects	✓	✓	✓
N	16,503	16,603	4,178
R^2	0.059	0.019	0.013

$^*p<0.05;$ $^{**}p<0.01;$ $^{***}p<0.001$

Table A3 – Characterization of the RDD effective sample. Predictors are explained in the paragraph above. HC2 heteroskedasticity consistent standard errors in parentheses.

A.7 Alternative outcome: Incumbent vote-share

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.031	-0.040*	-0.045	-0.055*
	0.022	0.023	0.029	0.031
Election cycle fixed effects	✓	✓	✓	✓
Controls		✓		✓
Bandwidth	0.365	0.332	0.365	0.332
N	1152	1047	1152	1047

*p<0.1; **p<0.05; ***p<0.01.

Table A4 – Effect of reaching the IDEB target on the vote-share of the incumbent. The bandwidth is the optimal bandwidth determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4.)
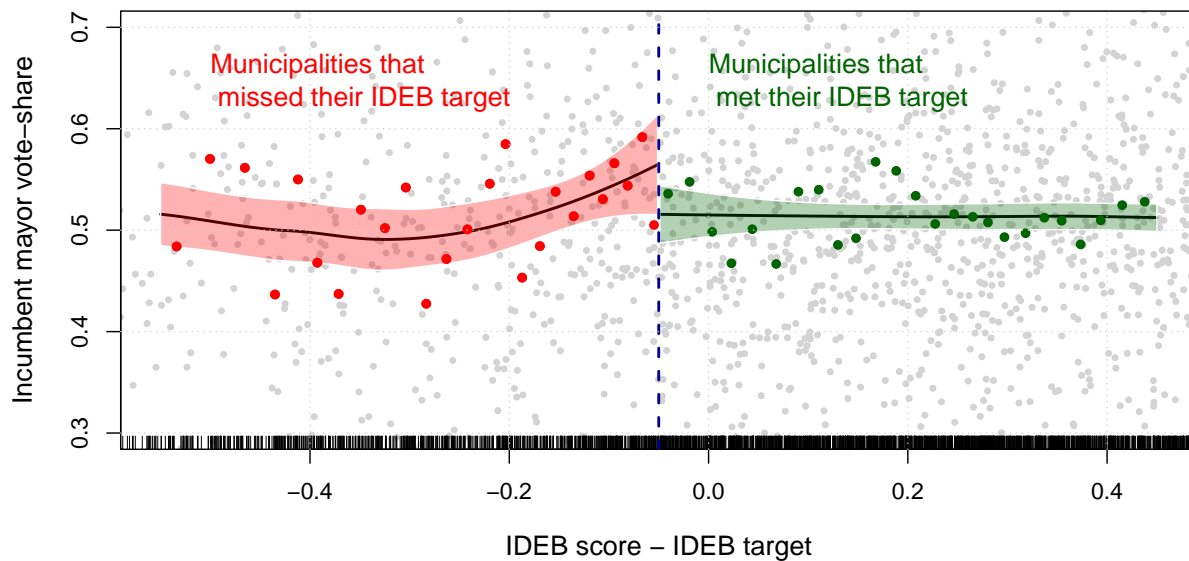
Figure A.5 – Relationship between meeting the IDEB target and vote share of the mayor. Grey dots are observations. Colored dots represent local averages for 50 equally-sized bins. Lines are loess regression lines estimated at both sides of the threshold with no controls. Shaded regions are their 95% confidence intervals.

RD results of the effect of meeting the IDEB on the vote-share of the incumbent are of similar standardized magnitude than those for re-election repoted in Table 1, but noisier. This is most likely due to the decrease in power stemming from a lower number of observations (municipalities where the mayor was allowed to run but chose not to are dropped from this sample but not that of Table 1). One may however be concerned that part of the difference may be driven by municipalities' IDEB performance driving incumbents' decisions to re-run. This is unlikely given the calendar for registering electoral candidates in Brazil: only in 2008 was this deadline *after* the release of IDEB scores.[12] Still, we test for this possibility by running the RD regression with whether the incumbent ran as a dependent variable. Table A5 shows statistically insignificant results.

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.038	-0.038	-0.042	-0.044
	0.040	0.035	0.050	0.052
Election cycle fixed effects	✓	✓	✓	✓
Controls		✓		✓
Bandwidth	0.470	0.446	0.470	0.446
N	2087	1998	2087	1998

$^{*}p<0.1; ^{**}p<0.05; ^{***}p<0.01.$

Table A5 – Effect of reaching the IDEB target on whether the mayor runs for re-election. The bandwidth is the optimal bandwidth determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4.)

---

[12]Candidates had to register their candidacies by July 7 in 2008, July 5 in 2012, and August 15 in 2016.

A.8 Alternative bandwidths
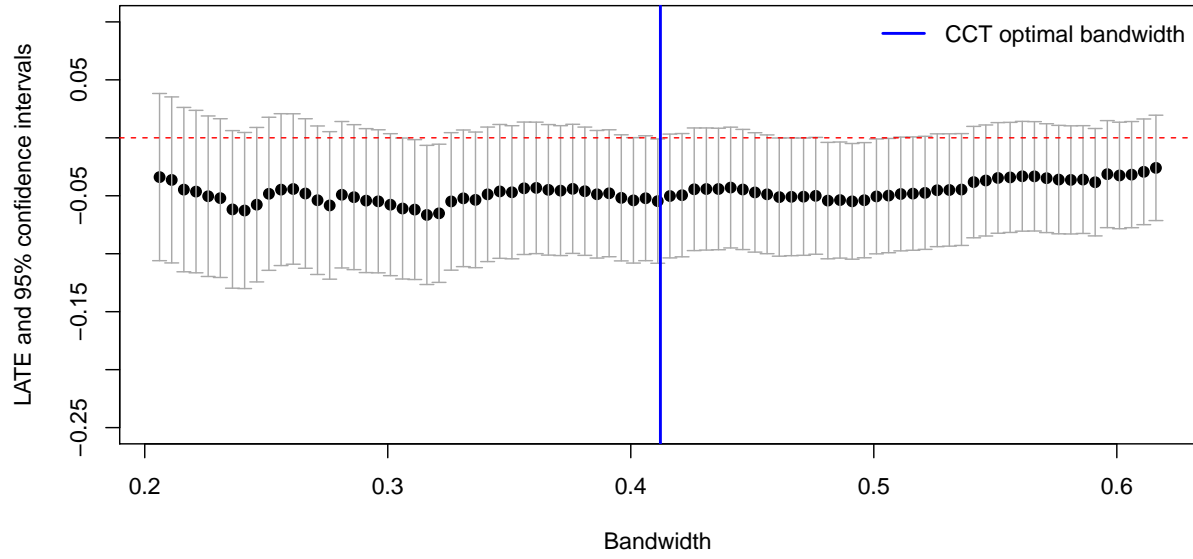
(a) Mayor re-election
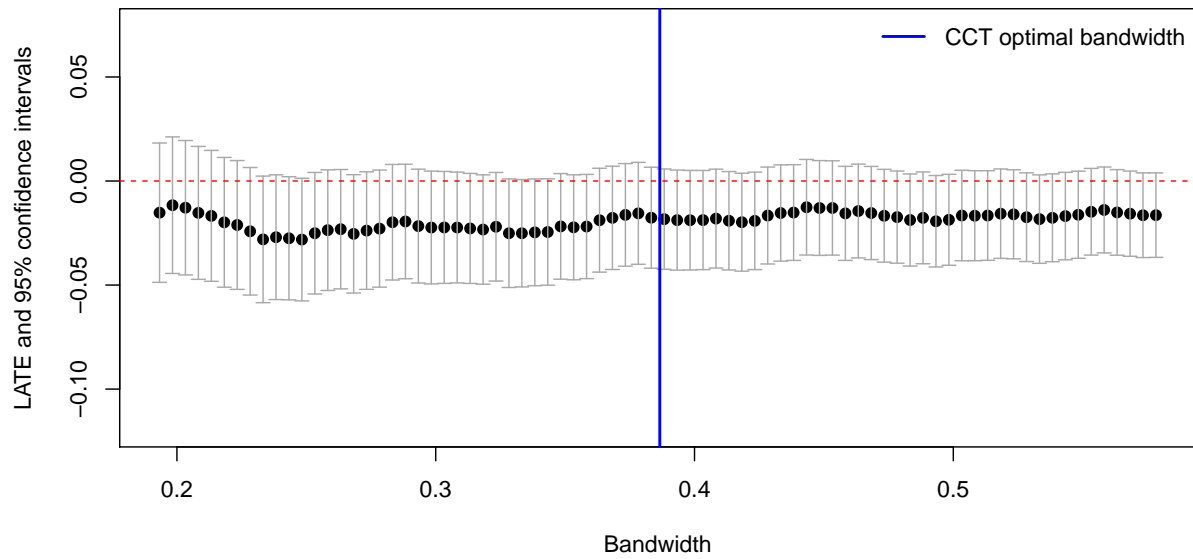
(b) Mayor vote-share

Figure A.6 – Robustness of the treatment effect shown in model 2 in Tables 1 and A4 to alternative bandwidths.

A.9   Alternative discontinuity thresholds as placebo tests
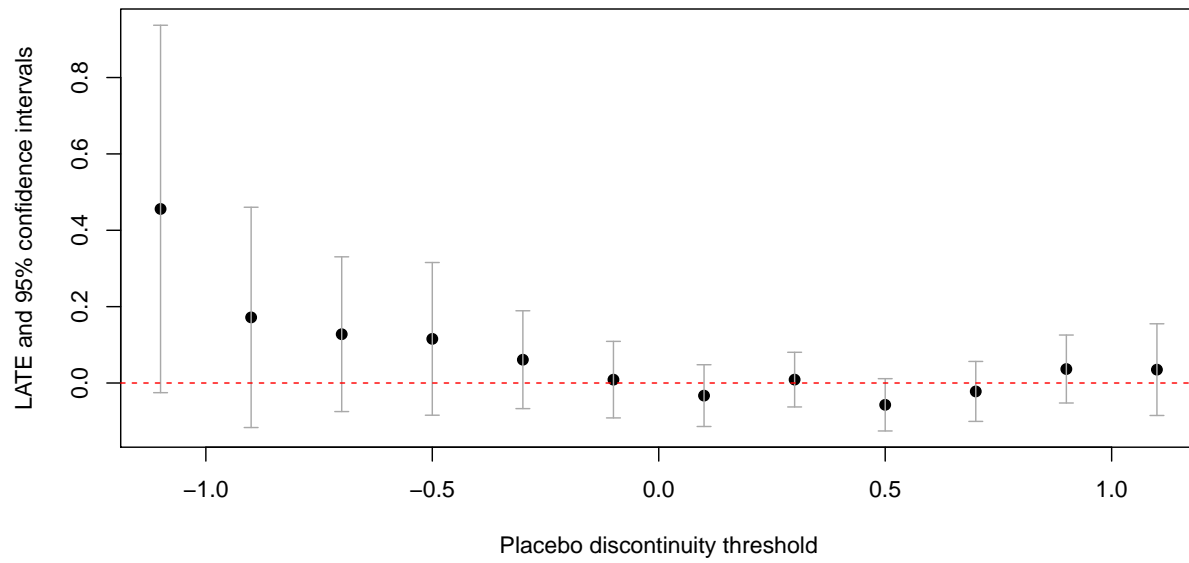
Figure A.7 – Placebo tests using alternative discontinuity thresholds for Model 2 in Table 1.

A.10 Alternative sample: no restrictions

As described in Section 4.1.1 and Appendix A6, our main results impose a number of constraints on the sample (observations where mayors are allowed to run, no supplementary elections are held, and municipalities receive only one IDEB signal). Here we present results lifting those constraints. Effects are smaller and less precisely estimated, but remain negative in all specifications.

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.023	-0.025	-0.021	-0.021
	0.022	0.021	0.028	0.029
Election cycle fixed effects	✓	✓	✓	✓
Covariates		✓		✓
Bandwidth	0.393	0.383	0.393	0.383
N	6293	6155	6293	6155

*p<0.1; **p<0.05; ***p<0.01.

c

Table A6 – Effect of reaching the IDEB target on the re-election of the incumbent. The bandwidth is the optimal bandwidth determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4.)

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.019*	-0.019	-0.024	-0.025
	0.011	0.012	0.015	0.016
Election cycle fixed effects	✓	✓	✓	✓
Covariates		✓		✓
Bandwidth	0.520	0.423	0.520	0.423
N	3853	3236	3853	3236

*p<0.1; **p<0.05; ***p<0.01.

Table A7 – Effect of reaching the IDEB target on the vote-share of the incumbent. The bandwidth is the optimal bandwidth determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4).

A.11 Alternative sample: wealthier municipalities

Bursztyn (2016) argues that the poor's preference for cash transfers (leading to immediate gains in consumption), depress demand for investments in education. To investigate whether the effects we find are driven by poor voters, we conduct the RDD analyses in the wealthier half of the municipalities (those where the share of residents who are poor is below the median). As shown below, effects are still negative in these municipalities—if less precisely estimated due to a substantial reduction in sample size.

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.054	-0.047	-0.086	-0.095
	0.054	0.051	0.066	0.068
Election cycle fixed effects	✓	✓	✓	✓
Covariates		✓		✓
Bandwidth	0.438	0.445	0.438	0.445
N	1282	1297	1282	1297

*p<0.1; **p<0.05; ***p<0.01.

Table A8 – Effect of reaching the IDEB target on the re-election of the incumbent. The bandwidth is the optimal bandwidth determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4.)

	Linear		Robust	
	(1)	(2)	(3)	(4)
IDEB target met	-0.045*	-0.042*	-0.057*	-0.060*
	0.025	0.025	0.033	0.035
Election cycle fixed effects	✓	✓	✓	✓
Covariates		✓		✓
Bandwidth	0.519	0.497	0.519	0.497
N	1017	977	1017	977

*p<0.1; **p<0.05; ***p<0.01.

Table A9 – Effect of reaching the IDEB target on the vote-share of the incumbent. The bandwidth is the optimal bandwidth determined by the algorithm of Calonico et al. (2014). Standard errors are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4).

B Randomized control trial

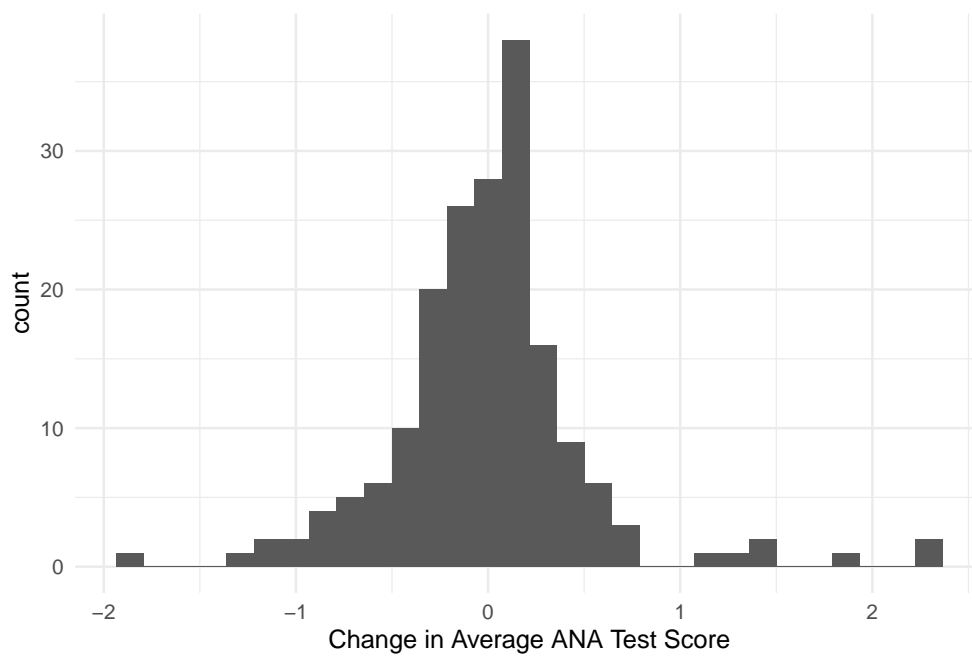## B.1 Distribution of ANA-based school quality scores

Figure B.8 – ANA distribution

B.2   Flyers Used in Experiment

Examples of the flyers used to deliver treatment information are contained in Figure B.9a (front side) and Figure B.9b (back side).
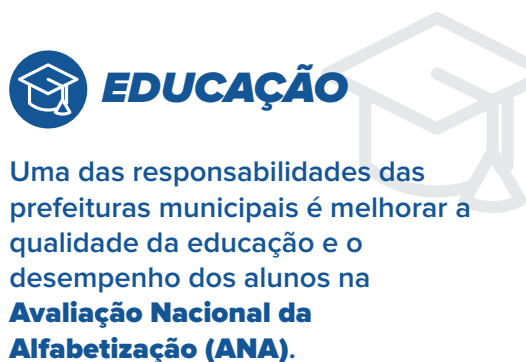
(a) Front of the flyer
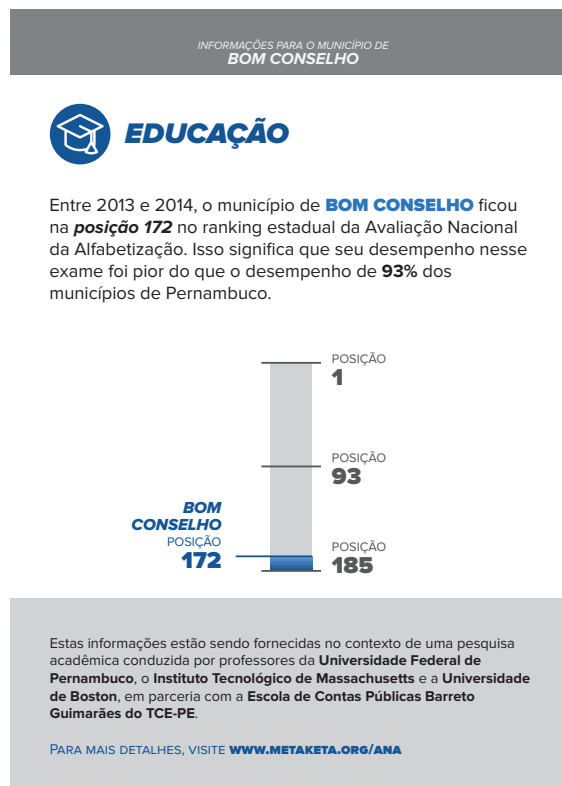
(b) Back of the flyer

Figure B.9 – Example of flyers Distributed to Voters

B.3 Legality of the Intervention

Some scholars have questioned the legality of conducting electoral field experiments in Brazil that do not involve partnerships with parties or candidates, given strict regulations governing campaign advertising (Cunow and Desposato, 2015; Desposato, 2015). Doubts about the legality of interventions involving flyers spring from Article 38 of Law 9504 of September 1997, which governs elections. With respect to "the dissemination of electoral advertising via the distribution of leaflets, stickers, flyers, and other printed material," it holds that such items "shall be published under the responsibility of the party, coalition, or candidate." However, this law does not precisely define what counts as "electoral advertising." Those sections that come closest to a definition suggest that, to be considered electoral advertising, a message must explicitly ask for votes. Article 26, paragraph 2 states that: "The following are considered electoral expenditures, subject to registry and to the limits set by this Law … direct and indirect advertising and publicity, via any medium of dissemination, intended to win votes." Likewise, Article 36-A, which governs campaigning prior to the official start date, holds that references to potential candidates "do not count as early electoral advertising as long as they do not explicitly ask for votes." Hence, a full reading of the suggests that flyers that do not mention voting are not subject to the limits of Article 38 because they do not count as electoral advertising.

To check our interpretation of Law 9504, we submitted a request for clarification to Brazil's Superior Electoral Court (TSE). Their response quoted Articles 26 and 38, as cited above, indicating that these were the relevant portions of the law bearing on the question of the legality of flyers. However, they told us that they could not provide any analysis or interpretation of the law, and that for that purpose we should contact a specialist attorney.

A condition of our partnership with the State Accounts Court of Pernambuco was that they would have the opportunity to review and approve all study materials before they went to the field. We submitted drafts of the flyers, which were reviewed by a TCE-PE Councilor who is a former judge of the Regional Electoral Court of Pernambuco (TRE-

PE), as well as a Substitute Councilor who is a law professor. They requested several changes to the draft version which we implemented prior to conducting the study.

Based on our inquiries with relevant legal authorities and specialists, we concluded that our intervention did not violate Brazilian law.

B.4 Example Ballot

An example of the secret ballot used to measure vote field experiment is contained in Figure B.10.

Figure B.10 – Secret Ballot for Measuring Vote Choice

PARA PREFEITO DE ABREU E LIMA

	NOME	NÚMERO	PARTIDO	
	KATIANA GADELHA	12	PDT	☐
	FLAVIO GADELHA	15	PMDB	☐
	PR. MARCOS JOSÉ	40	PSB	☐
🚫	BRANCO / NULO			☐

22

B.5 Covariates

As explained in the main text, we use a lasso procedure to select covariates to include in our regression adjustment. The pre-specified covariates are listed in Table A10. Most variables are originally likert-scale variables and we convert them to numeric interval variables. For categorical variables with many categories (`race`, `religion`, `muni_biggest_problem`), we collapse all categories with less than 5% of respondents into an "other" category and expand the variable to a set of dichotomous variables representing each possible value.

Variable	Definition
`female`	Female
`Age`	Age
`politics_interest`	Interest in politics
`turnout_2012`	Turnout in 2012 local elections
`vote_2012`	Voted for incumbent in 2012
`turnout_2014`	Turnout in 2014 national elections
`vote_2014`	Presidential candidate voter voted for in 2014
`partisan`	Identifies with a party
`muni_biggest_prob`	Most important problem in the municipality
`politician_helped`	Had received help from a politician
`govt_eval_baseline`	Evaluation of the municipal government at baseline
`acc_eval_baseline`	Evaluation of mayor's handling of the municipal accounts at baseline
`uncertain_acc_baseline`	Uncertainy over evaluation of handling of municipal accounts at baseline
`edu_eval_baseline`	Evaluation of mayor's handling of education
`uncertain_edu_baseline`	Uncertainty over evaluation of handling of education at baseline
`tce_knowledge`	Heard about the State Accounts Tribunal
`ana_knowledge`	Heard about ANA
`child_school`	Has a child in municipal school
`confid_fedgov`	Confidence in the federal government
`confid_justice`	Confidence in the Judiciary
`confid_tce`	Confidence in the State Accounts Tribunal
`confid_muni`	Confidence in the municipal government
`acc_responsible`	Degree to which mayor responsible for accounts status
`edu_responsible`	Degree to which mayor responsible for education
`prob_vote_buying`	Probability incumbent will attempt to buy vote
`prob_vote_monitoring`	Probability vote is not secret
`prob_vote_count`	Probability vote count is correct
`acc_rejected_prior`	Prior over whether municipal accounts are rejected
`tce_prior_cert`	Uncertainty about accounts status prior
`ana_prior`	Prior over ANA performance
`edu_prior_uncert`	Uncertainty over ANA prior
`years_edu`	Years of education
`race`	Race
`religion`	Religion
`income`	Income
`relative_wellbeing`	Perceived relative wellbeing
`edu_rank1`	Ranks handling of education as more important than handling of accounts.
`acc_rank1`	Ranks handling of account as more important than education.

Table A10 – Covariates for the Pernambuco Field Experiment.

B.6 Experimental Results without Covariate Adjustment

	All	Parents	Not Parents	All
Treatment	-0.0317	-0.0489	-0.0145	-0.0128
	(0.0229)	(0.0529)	(0.0286)	(0.0279)
Treatment x Rank	0.0003	-0.0013	0.0008	0.0009
	(0.0004)	(0.0009)	(0.0005)	(0.0005)
Treatment x Rank x Parents				-0.0020^*
				(0.0009)
Num. obs.	1709	525	1184	1709

$^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$

Table A11 – Experimental results without covariate adjustment. Experimental block coefficients are omitted. HC2 heteroskedasticity consistent standard errors in parentheses.

Results from the experiment without any covariate adjustment (aside from adjustment for experimental blocks) are presented in Table A11.

B.7 Robustness to Other Interactions

	All	Parents	Not Parents
Treatment	-0.0553^{**}	-0.0704^{*}	-0.0422
	(0.0196)	(0.0352)	(0.0239)
Treatment x Rank	0.0004	-0.0007	0.0008^{*}
	(0.0003)	(0.0006)	(0.0004)
Num. obs.	1709	525	1184

$^{***}p < 0.001, {}^{**}p < 0.01, {}^{*}p < 0.05$

Table A12 – Experimental results with additional interactions. These estimates are from the regression models reported in Table 2 augmented with the interaction between the treatment and age, income, and years of education. HC2 heteroskedasticity consistent standard errors in parentheses.

In the discussion of the contrasting results of the experimental intervention on parents of children enrolled in municipal schools and non-parents, we note that parents tend to be younger, poorer, and less educated than non-parents. Given these differences, we control for these covariates by including them as additional interactions in the first three specifications found in Table 2. The findings that parents punish bad performers and that non-parents tend to punish good performers are robust to including these covariates as additional interactions, as shown in Table A12. The point estimates in Table A12 reported on the interaction between the intervention and performance on the ANA are almost identical to those reported in Table 2.

B.8 Heterogeneity by Priors

The result in Figure 2 might be explained by a correlation between rank and voter expectations. If voters in higher performing municipalities expected even better performance than the actual rank, while those in lower performing municipalities had properly calibrated expectations, then we would expect a negative interaction like the one we found. To test for this possibility, we estimate heterogeneity by the *gap* between voters' prior belief over their municipality's rank (measured at baseline) and the actual rank.

Results for treatment heterogeneity by the gap between voters expectations and reality can be found in Figure B.11. As is clearly evident from the plotted regression line, there is no substantive interaction between this gap and the effect of the treatment. The overall effect of treatment is negative, but this negative effect does not moderate among
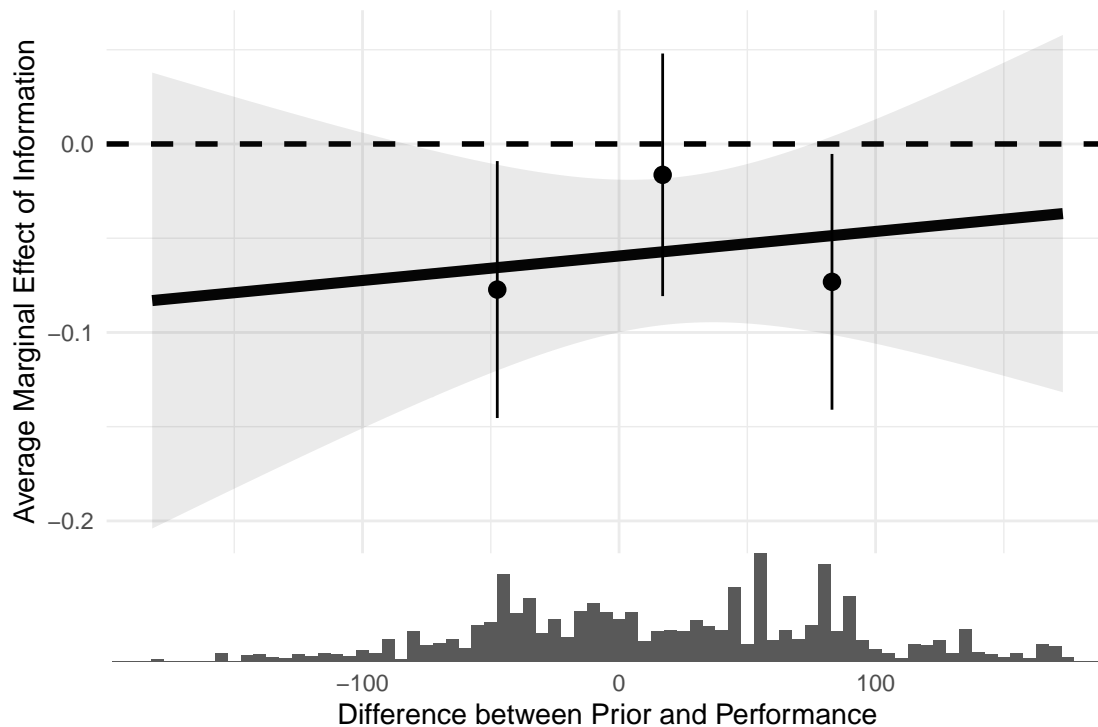
Figure B.11 – Effect of Treatment by Gap Between Voters' Prior and Municipal Performance. Negative values indicate worse performance than expected, while positive values indicate better performance than expected. 90% Confidence Intervals are shown. Histogram shows marginal distribution of the data by the gap.

voters who are positively surprised by the information. The tercile approach largely corroborates the linear interaction model, though there is some weak indication of a non-linearity with voters in the middle of the distribution punishing the incumbent slightly more than incumbents at the tails of the distribution.

B.9 Experiment Covariate Balance

Variable	Mean Difference	SD	SE	Permutation p-value
partisan	0.0626	0.49	0.0215	0.0095
turnout_2012	-0.0413	0.353	0.0167	0.0255
acc_rank1	-0.0349	0.435	0.0188	0.067
acc_eval_baseline	0.103	1.31	0.0536	0.073
politician_helped	0.0266	0.402	0.0184	0.177
confid_justice	0.121	2.08	0.093	0.198
tce_knowledge	-0.0285	0.497	0.0217	0.207
confid_muni	0.12	2.3	0.0995	0.252
edu_rank1	-0.0217	0.434	0.019	0.279
politics_interest	0.0477	0.962	0.042	0.295
confid_tce	0.0967	1.96	0.0935	0.339
turnout_2014	-0.0142	0.302	0.0141	0.359
acc_rejected_prior	-0.0203	0.44	0.0193	0.305
confid_fedgov	0.0957	2.1	0.0933	0.336
edu_eval_baseline	0.0571	1.26	0.0516	0.289
govt_eval_baseline	-0.0596	1.38	0.0579	0.329
uncertain_acc_baseline	-0.0418	1.29	0.0506	0.444
vote_2012	0.0155	0.484	0.0206	0.456
edu_responsible	-0.0504	1.61	0.055	0.388
tce_prior_cert	-0.018	0.598	0.0248	0.511
child_school	-0.013	0.461	0.0209	0.573
edu_prior_uncert	0.0157	0.567	0.0221	0.497
prob_vote_monitoring	-0.0277	1.15	0.0495	0.596
age	0.361	16.3	0.79	0.678
income	0.0214	1.46	0.0627	0.749
years_edu	-0.069	4.82	0.215	0.76
relative_wellbeing	0.00827	0.595	0.0268	0.773
ana_knowledge	0.00475	0.426	0.0189	0.806
female	-0.00524	0.5	0.0242	0.843
ana_prior	-0.506	51.8	2.52	0.856
prob_vote_count	0.0075	1.13	0.0407	0.862
prob_vote_count	0.0075	1.13	0.0407	0.872
acc_responsible	0.00844	1.56	0.0539	0.878
uncertain_edu_baseline	-0.00465	1.35	0.0538	0.93

Table A13 – Covariate Balance for the Pernambuco Experiment. The mean difference column reports estimated ATE for pre-treatment covariates specified in PAP. SD is standard deviation of variable. Permutation p-value is computed using 2000 permutations of the treatment variable. Most imbalanced variables are at the top of the table, as variables are ordered by the absolute value of the mean difference divided by the standard deviation.

Covariate balance is presented in Table A13. All variables in Table A10 are included,

except for the non-ordinal categorical variables (`race`, `vote_2014`, `religion`, `muni_biggest_prob`). *p*-values calculated using permutation inference with 2000 permuted treatment realizations.

B.10 Results for Pre-Registered Hypotheses

As mentioned in the main text, our analytical specification for the field experiment departs from that which was pre-registered with Evidence in Governance and Politics (EGAP) prior to analysis. In concert with the broader initiative of which this study was a part, our pre-specified approach conditions on whether information is "good news" or "bad news" (measured dichotomously) relative to a respondent's priors. By contrast, in the main text, we ignore priors and condition only on the municipality's continuous performance ranking.

For the treatment arm involving information about school performance, we pre-specified 23 distinct hypothesis tests. Results from these tests can be found in Table A14. As shown in line 1 of the table, our main overall finding from the paper is similar when using the pre-specified approach. Information that is "good news" relative to priors reduces the probability of voting for the incumbent by 0.083, significant at the 0.05 level. By comparison, for respondents in the best tercile of ANA performance (ignoring priors), treatment information has an effect of -0.073, similarly significant at the 0.05 level. Hence, regardless of whether we condition on priors, respondents punish good performance.

We refer readers to the pre-analysis plan for complete details, but will describe here the main variables and subsets of the data used in each specification. The subset "Good News" refers to individuals who received information that is more positive than their priors (or equal to priors and better than or equal to the statewide median), while "Bad News" refers to individuals who received information that is worse than their priors (or equal to priors and worse than the statewide median). "All" refers to all individuals regardless of the content of information received.

The outcome variables are as defined as follows:

- `vote`: Equals 1 if the respondent reports voting for the incumbent, 0 otherwise (including abstention and blank/null votes).

- `turnout`: Equals 1 if the respondent reports voting in the 2016 election, 0 otherwise.

31

Family	Hypothesis	Outcome	Parameter	Subset	Estimate	SE	p-value
1	1a	vote	ana	Good News	-0.083	0.033	0.013
1	1b	vote	ana	Bad News	-0.041	0.028	0.137
2	2	vote	ana x uncertain_edu	Good News	0.001	0.001	0.177
2	2	vote	ana x uncertain_edu	Bad News	0.000	0.001	0.445
2	3	vote	ana x edu_rank1	Good News	-0.094	0.074	0.202
2	3	vote	ana x edu_rank1	Bad News	-0.061	0.066	0.355
2	4	vote	ana x edu_rank1 x uncertain_edu	Good News	-0.002	0.050	0.964
2	4	vote	ana x edu_rank1 x uncertain_edu	Bad News	-0.040	0.052	0.443
2	5a	vote	ana x child_school	Good News	-0.030	0.071	0.678
2	5a	vote	ana x child_school	Bad News	-0.067	0.061	0.274
2	5b	vote	ana x child_school x uncertain_edu	Good News	0.074	0.051	0.148
2	5b	vote	ana x child_school x uncertain_edu	Bad News	0.086	0.045	0.057
3	7a	turnout	ana	Good News	-0.714	0.577	0.217
3	7b	turnout	ana	Bad News	0.021	0.013	0.105
4	8a	valid_vote	ana	Bad News	0.027	0.020	0.185
4	8b	valid_vote	ana x govt_eval	Bad News	0.045	0.036	0.207
4	9	edu_eval	ana	Good News	-0.009	0.087	0.918
4	9	edu_eval	ana	Bad News	-0.059	NaN	NaN
4	10	edu_eval	ana x uncertain_edu	Good News	-0.081	0.057	0.157
4	10	edu_eval	ana x uncertain_edu	Bad News	0.013	0.054	0.805
4	11	uncertain_edu_eval	ana	All	-0.017	0.032	0.592
4	12	info_import	ana	All	-0.047	0.028	0.098
4	13	ana_correct	ana	All	4.244	1.954	0.030

Table A14 – Results for Pre-Registered Hypotheses. See pre-analysis plan for details.

- valid_vote: Using a secret ballot, equals 1 if the respondent reports having voted for any candidate for mayor, and 0 if s/he reports having abstained or cast a blank or null vote.

- edu_eval: The respondent's evaluation of the incumbent's management of schools at endline: (1) "excellent," (2) "good," (3) "regular," (4) "bad," or (5) "horrible."

- uncertain_edu_eval: The respondent's certainty about edu_eval: "very sure," (2) "sure," (3) "unsure," or (4) "very unsure."

- info_import: Equals 1 if the respondent said that "what I learned during the campaign about the mayor" was one of the three most important factors in their decision-making; 0 otherwise.

- ana_correct: The degree to which the respondent's post-treatment belief about the municipality's ANA ranking is accurate: 184 - |ana_posterior - ana_rank|.

  The independent and moderator variables are defined as follows:

- ana: Equals 1 for respondents assigned to receive information on municipal performance on the ANA, 0 otherwise.

- `uncertain_edu`: The respondent's certainty about the respondent's own evaluation of the incumbent's management of schools. Takes on integer values from 1 ("totally sure") to 7 ("not at all sure").

- `edu_rank1`: Equals 1 if respondent ranks education above financial management in importance, 0 otherwise.

- `child_school`: Equals 1 if the respondent has a child in municipal schools, 0 otherwise.

- `govt_eval`: The respondent's evaluation of the incumbent's administration at endline: (1) "excel- lent," (2) "good," (3) "regular," (4) "bad," or (5) "horrible."

C   Online survey experiment

C.1   Respondent recruitment via Facebook ads

We recruited respondents for the online survey experiment via Facebook ads, a well-established, low-cost method for subject recruitment in comparative politics (Boas et al., 2018). Targeted Facebook users saw advertisements offering a chance to win 100 Brazilian reais (about 24 US dollars) for taking a survey, with a 1/100 probability of winning. To ensure a diverse sample, ads were targeted to respondents in distinct strata of region and age. We also excluded Brazil's 300 largest municipalities from our advertising campaign in order to recruit enough respondents from small and medium-sized municipalities, which make up the immense majority of the sample in the RDD and field experiment. Finally, to increase statistical power, we more heavily targeted states where a larger share of municipalities did not meet their IDEB target, which is the less common outcome. We aimed to recruit 3,000 Brazilians, and ended up surveying 3,019 respondents.

C.2   Balance in covariates

We examine balance in pre-treatment covariates by regressing each of them on a general treatment indicator, and then subsetting to subjects who were or would have been exposed to positive vs negative information, in line with our analyses. Table A15 presents the results. We observe some imbalance in priority given to education (for the overall treatment) and priority given to public jobs (for general and positive treatment). Note however that these two variables are dependent on each other, and imbalance in the rank given to one policy area can be linked mechanically to imbalance in the rank given to another policy area. This is especially true since, as shown in Appendix C3, jobs and education tend to get ranked lowly and highly, respectively. The dummy version of priority given to education, which we pre-registered as a hypothetical mediator, is balanced across groups. We also observe some imbalance for age in the positive treatment group. All in all, about 6% of our covariates show a statistically significant difference (at the 95% level) between treatment and control groups.

	Treatment			Positive treatment			Negative treatment		
	$\bar{T} - \bar{C}$	SE	p	$\bar{T} - \bar{C}$	SE	p	$\bar{T} - \bar{C}$	SE	p
Measured before treatment									
Region = North	0.012	0.010	0.247	0.016	0.012	0.201	0.003	0.018	0.881
Region = Center-West	0.011	0.009	0.225	0.011	0.011	0.336	0.013	0.016	0.395
Region = Southeast	-0.027	0.017	0.109	-0.027	0.020	0.173	-0.022	0.030	0.467
Region = South	-0.001	0.014	0.930	-0.007	0.016	0.662	0.014	0.026	0.600
Region = Northeast	0.005	0.017	0.780	0.008	0.020	0.706	-0.008	0.034	0.819
Priority given to education	-0.106	0.049	0.031	-0.100	0.057	0.082	-0.123	0.094	0.189
Priority education = Low	-0.020	0.018	0.275	-0.021	0.021	0.325	-0.017	0.034	0.613
Priority given to healthcare	0.001	0.047	0.976	-0.010	0.055	0.855	0.031	0.088	0.721
Priority given to anti-poverty	-0.003	0.055	0.961	-0.019	0.065	0.771	0.037	0.105	0.727
Priority given to economy in general	-0.050	0.052	0.343	-0.054	0.062	0.386	-0.039	0.098	0.695
Priority given to public jobs	0.113	0.048	0.020	0.125	0.057	0.029	0.082	0.091	0.369
Priority given to security	0.044	0.051	0.385	0.057	0.059	0.335	0.012	0.097	0.902
Had heard of IDEB before	-0.023	0.017	0.158	-0.036	0.019	0.063	0.011	0.032	0.733
Measured after treatment									
Age	-0.856	0.456	0.061	-1.381	0.531	0.009	0.521	0.884	0.556
Female	-0.004	0.017	0.807	0.013	0.020	0.525	-0.047	0.033	0.152
Education	0.030	0.085	0.724	-0.039	0.100	0.695	0.213	0.161	0.185
Race = white	-0.029	0.018	0.110	-0.034	0.021	0.112	-0.014	0.034	0.682
Race = brown	0.018	0.018	0.312	0.017	0.021	0.426	0.020	0.034	0.569
Race = other	0.011	0.013	0.401	0.017	0.015	0.255	-0.006	0.024	0.814
Has kids	0.007	0.017	0.681	0.011	0.020	0.581	-0.000	0.033	1.000
Has kids in basic education	0.000	0.015	0.992	-0.002	0.018	0.910	0.007	0.028	0.795
Has kids in municipal school	0.016	0.013	0.205	0.012	0.015	0.442	0.028	0.023	0.222

Table A15 – Difference in means across treatment groups
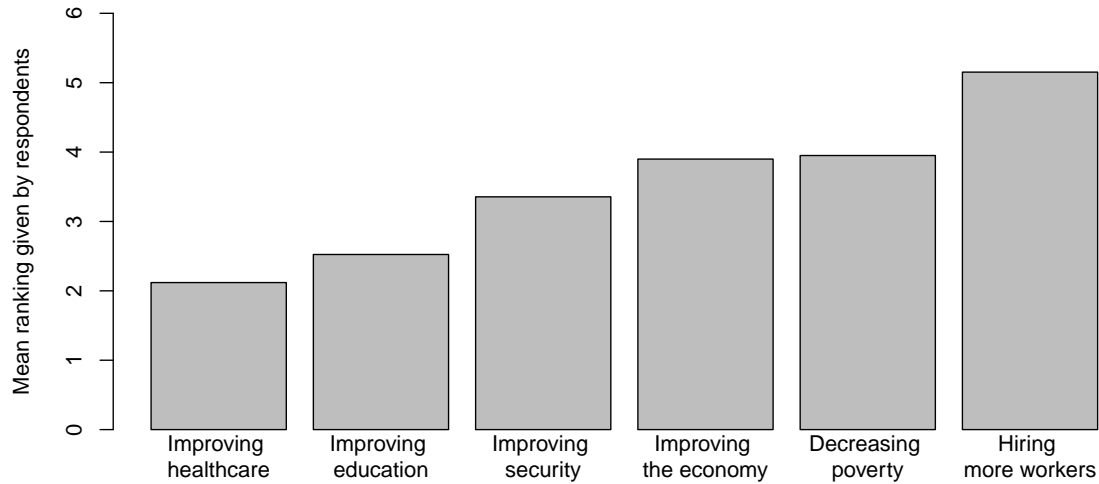
C.3   Ranking of policy areas by respondents

Figure C.12 – Mean ranking given by online survey respondents when asked to rank policy areas according to the priority they should be given by their municipal government

C.4 Additional results

The pages below present the regression results corresponding to all specifications in the pre-analysis plan for the online survey experiment.

Table A16 – Online survey experiment results: Splitting sample by whether respondent is from a municipality where IDEB was met

	I would vote for them (1)	They hired many workers (2)	They improved clinics (3)	They reduced poverty (4)	They improved schools (5)	They invested a lot in educ. (6)	They helped people like me (7)
	Dependent variable: 'How much do you agree with the following statements about the mayor?' on a 1-4 scale (1 = disagree completely, 4 = agree completely)						
$Tr.^{+}$	0.095	0.101*	0.093*	0.101*	0.123**	0.110*	0.075
	(0.051)	(0.044)	(0.046)	(0.043)	(0.046)	(0.045)	(0.046)
Constant	2.208***	2.607***	2.512***	1.960***	2.476***	2.290***	2.103***
	(0.036)	(0.031)	(0.033)	(0.030)	(0.032)	(0.032)	(0.033)
N	2,173	2,173	2,173	2,173	2,173	2,173	2,173
R^2	0.021	0.017	0.020	0.023	0.017	0.017	0.015
$Tr.^{-}$	−0.230**	−0.029	−0.189*	−0.201**	−0.086	−0.204**	−0.198*
	(0.082)	(0.073)	(0.077)	(0.070)	(0.074)	(0.072)	(0.077)
Constant	2.224***	2.603***	2.554***	2.018***	2.402***	2.253***	2.165***
	(0.059)	(0.052)	(0.055)	(0.050)	(0.053)	(0.051)	(0.055)
N	846	846	846	846	846	846	846
R^2	0.038	0.041	0.040	0.033	0.054	0.051	0.028

All regressions control for age, gender, education, race, and region, as specified in Equation 5
HC2 standard errors in brackets
*$p<0.05$; **$p<0.01$; ***$p<0.001$

Table A17 – Online survey experiment results: Respondents from all municipalities

	Dependent variable: 'How much do you agree with the following statements about the mayor?' *on a 1-4 scale (1 = disagree completely, 4 = agree completely)*						
	I would vote for them	They hired many workers	They improved clinics	They reduced poverty	They improved schools	They invested a lot in educ.	They helped people like me
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Tr.^+$	0.089	0.103*	0.081	0.082*	0.141***	0.120**	0.059
	(0.048)	(0.041)	(0.043)	(0.040)	(0.042)	(0.042)	(0.043)
$Tr.^-$	−0.220***	−0.031	−0.159**	−0.161**	−0.142*	−0.231***	−0.151*
	(0.066)	(0.057)	(0.060)	(0.055)	(0.058)	(0.058)	(0.060)
Constant	2.214***	2.605***	2.525***	1.978***	2.458***	2.280***	2.119***
	(0.031)	(0.026)	(0.028)	(0.026)	(0.027)	(0.027)	(0.028)
N	3,019	3,019	3,019	3,019	3,019	3,019	3,019
R^2	0.023	0.019	0.022	0.021	0.023	0.026	0.016

All regressions control for age, gender, education, race, and region, as specified in Equation **??**
HC2 standard errors in brackets
*$p<0.05$; **$p<0.01$; ***$p<0.001$

Table A18 – Online survey experiment results: Respondents from municipalities where the school quality target was met, with treatment heterogeneity by priority given to education

	Dependent variable: 'How much do you agree with the following statements about the mayor?' *on a 1-4 scale (1 = disagree completely, 4 = agree completely)*						
	I would vote for them	They hired many workers	They improved clinics	They reduced poverty	They improved schools	They invested a lot in educ.	They helped people like me
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Tr.^{+}$	0.140**	0.125**	0.202***	0.167***	0.203***	0.151**	0.129**
	(0.068)	(0.058)	(0.061)	(0.057)	(0.060)	(0.060)	(0.061)
$Tr.^{+} \times EL$	−0.133	−0.058	−0.265***	−0.149*	−0.171*	−0.092	−0.108
	(0.105)	(0.090)	(0.095)	(0.089)	(0.093)	(0.093)	(0.095)
EL	0.165**	0.058	0.236***	0.195***	0.171***	0.187***	0.197***
	(0.074)	(0.064)	(0.067)	(0.062)	(0.066)	(0.065)	(0.067)
Constant	2.143***	2.575***	2.412***	1.870***	2.391***	2.206***	2.007***
	(0.048)	(0.041)	(0.043)	(0.040)	(0.043)	(0.042)	(0.043)
N	2,173	2,173	2,173	2,173	2,173	2,173	2,173
R^2	0.033	0.029	0.034	0.035	0.032	0.030	0.027

EL stands for "education low", an indicator for whether the respondent gives below-median priority to education.
All regressions control for age, gender, education, race, and region, as specified in Equation **??**
HC2 standard errors in brackets
*$p<0.05$; **$p<0.01$; ***$p<0.001$

Table A19 – Online survey experiment results: Respondents from municipalities where the school quality target was not met, with treatment heterogeneity by priority given to education

	Dependent variable: 'How much do you agree with the following statements about the mayor?' on a 1-4 scale (1 = disagree completely, 4 = agree completely)						
	I would vote for them	They hired many workers	They improved clinics	They reduced poverty	They improved schools	They invested a lot in educ.	They helped people like me
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Tr.^-$	−0.200*	−0.023	−0.194*	−0.109	−0.011	−0.123	−0.145
	(0.113)	(0.100)	(0.106)	(0.095)	(0.101)	(0.098)	(0.106)
$Tr.^- \times EL$	−0.059	−0.038	0.026	−0.185	−0.154	−0.143	−0.146
	(0.172)	(0.152)	(0.162)	(0.145)	(0.155)	(0.150)	(0.162)
EL	0.019	0.097	−0.031	0.223**	0.045	0.176	0.104
	(0.123)	(0.109)	(0.116)	(0.104)	(0.111)	(0.107)	(0.116)
Constant	2.225***	2.563***	2.567***	1.927***	2.387***	2.181***	2.135***
	(0.082)	(0.073)	(0.077)	(0.069)	(0.074)	(0.072)	(0.077)
N	846	846	846	846	846	846	846
R^2	0.060	0.066	0.053	0.066	0.075	0.074	0.044

EL stands for "education low", an indicator for whether the respondent gives below-median priority to education.
All regressions control for age, gender, education, race, and region, as specified in Equation **??**
HC2 standard errors in brackets
*p<0.05; **p<0.01; ***p<0.001

Table A20 – Online survey experiment results: Respondents from municipalities where the school quality target was met, with treatment heterogeneity by whether respondents have children in municipal schools

	Dependent variable: 'How much do you agree with the following statements about the mayor?' on a 1-4 scale (1 = disagree completely, 4 = agree completely)						
	I would vote for them	They hired many workers	They improved clinics	They reduced poverty	They improved schools	They invested a lot in educ.	They helped people like me
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Tr.^+$	0.087	0.111*	0.089	0.114*	0.115*	0.129**	0.058
	(0.056)	(0.048)	(0.050)	(0.047)	(0.049)	(0.049)	(0.050)
$Tr.^+ \times$ CS	0.018	−0.023	0.025	−0.128	0.042	−0.125	0.175
	(0.153)	(0.132)	(0.138)	(0.129)	(0.136)	(0.135)	(0.138)
CS	0.154	0.050	0.157	0.237**	0.130	0.240*	0.065
	(0.109)	(0.093)	(0.098)	(0.091)	(0.096)	(0.096)	(0.098)
Constant	2.193***	2.597***	2.491***	1.931***	2.461***	2.254***	2.090***
	(0.039)	(0.033)	(0.035)	(0.033)	(0.035)	(0.034)	(0.035)
N	2,173	2,173	2,173	2,173	2,173	2,173	2,173
R^2	0.031	0.023	0.033	0.033	0.028	0.029	0.025

CS stands for "children in school", an indicator for whether respondents have children enrolled in municipal schools.
All regressions control for age, gender, education, race, and region, as specified in Equation **??**
HC2 standard errors in brackets
*p<0.05; **p<0.01; ***p<0.001

Table A21 – Online survey experiment results: Respondents from municipalities where the school quality target was not met, with treatment heterogeneity by whether respondents have children in municipal schools

	Dependent variable: 'How much do you agree with the following statements about the mayor?' on a 1-4 scale (1 = disagree completely, 4 = agree completely)						
	I would vote for them	They hired many workers	They improved clinics	They reduced poverty	They improved schools	They invested a lot in educ.	They helped people like me
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Tr.^-$	−0.211*	0.001	−0.196*	−0.171*	−0.084	−0.213**	−0.213**
	(0.088)	(0.078)	(0.083)	(0.075)	(0.080)	(0.077)	(0.082)
$Tr.^- \times CS$	−0.174	−0.224	0.022	−0.160	−0.099	0.126	0.180
	(0.280)	(0.248)	(0.261)	(0.236)	(0.252)	(0.243)	(0.261)
CS	0.108	0.335	0.238	0.153	0.173	0.150	0.123
	(0.210)	(0.187)	(0.196)	(0.178)	(0.189)	(0.183)	(0.196)
Constant	2.200***	2.568***	2.516***	1.990***	2.375***	2.220***	2.131***
	(0.063)	(0.056)	(0.058)	(0.053)	(0.056)	(0.054)	(0.058)
N	846	846	846	846	846	846	846
R^2	0.059	0.059	0.068	0.064	0.072	0.082	0.063

CS stands for "children in school", an indicator for whether respondents have children enrolled in municipal schools.
All regressions control for age, gender, education, race, and region, as specified in Equation **??**
HC2 standard errors in brackets
*p<0.05; **p<0.01; ***p<0.001

Table A22 – Online survey experiment results: Index of mayor performance on non-education areas

	Dependent variable: Index of mayor performance on non-education items						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Tr.^+$	0.149*	0.173**		0.296***		0.173*	
	(0.063)	(0.067)		(0.088)		(0.072)	
$Tr.^-$	−0.251**		−0.309**		−0.234		−0.293*
	(0.086)		(0.111)		(0.153)		(0.119)
$Tr.^+ \times$ EL				−0.283*			
				(0.137)			
$Tr.^+ \times$ CS						0.022	
						(0.199)	
$Tr.^- \times$ EL					−0.174		
					(0.234)		
$Tr.^- \times$ CS							−0.064
							(0.376)
EL				0.338***	0.191		
				(0.096)	(0.167)		
CS						0.254	0.377
						(0.141)	(0.283)
Constant	−0.013	−0.037	0.045	−0.192**	−0.027	−0.072	−0.018
	(0.040)	(0.047)	(0.080)	(0.062)	(0.112)	(0.051)	(0.084)
Observations	3,019	2,173	846	2,173	846	2,173	846
R^2	0.021	0.022	0.035	0.038	0.053	0.034	0.068

Dependent variable is the first component from running PCA on respondents'
assessment of mayor performance in all non-education areas.
EL is an indicator for whether the respondent gives below-median priority to education.
CS is an indicator for whether respondents have children enrolled in municipal schools.
All regressions control for age, gender, education, race, and region, as specified in Equation **??**
HC2 standard errors in brackets
*p<0.05; **p<0.01; ***p<0.001

D   Null relationship between municipal education spending and school quality scores

To examine whether there is a relationship between municipal education spending and municipal primary school quality scores, we use official data from Brazil's National Treasury (FINBRA) and from the Ministry of Education. We first calculate the relative increase in spending and IDEB scores between 2013 and 2017 (dividing the increase by the baseline level), and regress the first on the second:

The resulting slope coefficient measures the association between spending in education and school quality indicators. By using a first-differenced model, we get rid of any time-invariant confounding. Results are presented in Table A23 below. There is no statistically significant association between increases in spending and any of the measures of school quality improvement that we consider: whether the municipality met its IDEB target in 2017, the increase of IDEB scores in the period, the increase in the gap between the IDEB score and the IDEB target in the period, and the increase in IDEB scores relative to scores at baseline. The latter is our preferred specification, since it accounts for baseline levels of IDEB. Figure D.13 presents this null relationship from Model 4. All in all, the data suggest increases in spending are not correlated with increases in school quality.

Menezes Filho et al. (2009) reach a similar conclusion after analyzing the relationship between municipal spending in education school performance in 2005. Monteiro (2015) finds a null relationship between increased education spending and test scores, leveraging as an instrumental variable increases in municipal revenue stemming from oil royalties.

	IDEB target met in 2017	Δ 2013 - 2017 in IDEB:		
		Score	Score - Target	Score, rel. to 2013
	(1)	(2)	(3)	(4)
Δ 2013 - 2017 in education spending, relative to 2013	0.082 (0.064)	0.022 (0.064)	0.026 (0.051)	0.00001 (0.016)
N	4,886	4,653	4,632	4,653
R^2	0.001	0.00003	0.00004	0.000

*p<0.05; **p<0.01; ***p<0.001

Table A23 – Relationship between municipal spending in primary education and performance of municipal primary education schools in IDEB. HC2 heteroskedasticity consistent standard errors in parentheses. Intercept omitted from the table.
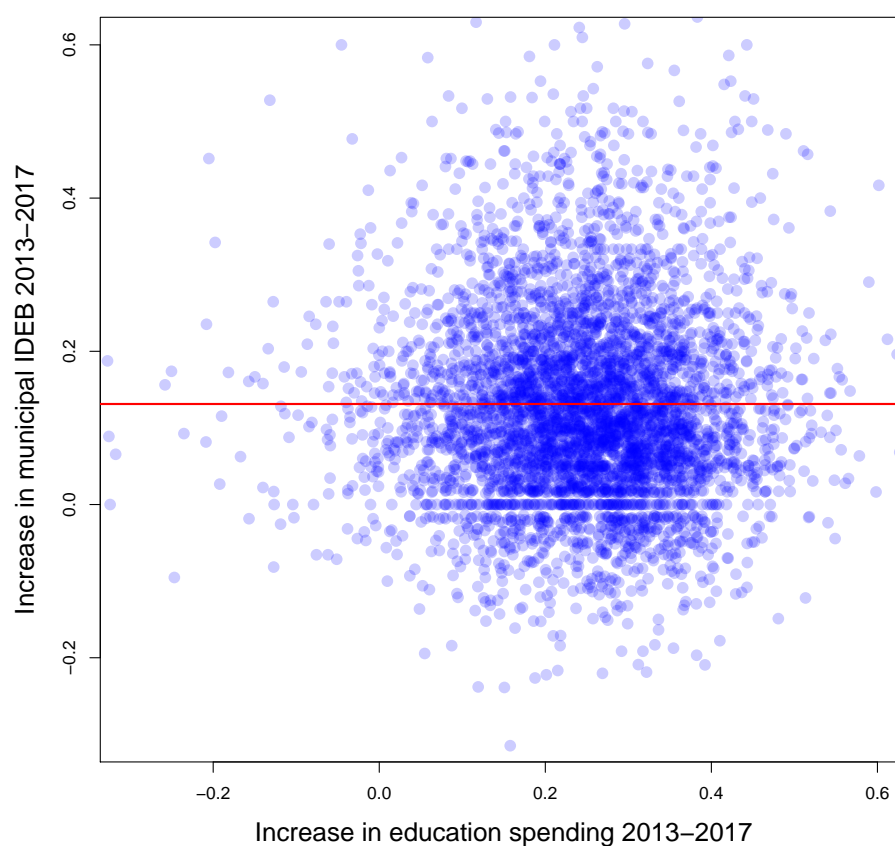

Figure D.13 – Relationship between relative increases in municipal education spending and relative increases in municipal school quality scores. Red line is the regression line from Model 4 in Table A23.