Competence versus Priorities: Negative Electoral Responses to Education Quality in Brazil

Taylor Boas
Boston University
tboas@bu.edu

F. Daniel Hidalgo
MIT
dhidalgo@mit.edu

Guillermo Toral
MIT
gtoral@mit.edu

August 21, 2020

Forthcoming, *Journal of Politics*

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 a regression discontinuity design 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, we argue that voters who do not value education interpret school quality as an indicator of municipal policy priorities and perceive trade-offs with other services. Voters may hold politicians accountable not only for their competence but also for their representation of potentially conflicting interests—a fact that complicates the simple logic behind many accountability interventions.

Keywords: Information; Elections; Accountability; Education

Supplementary material for this article is available in the appendix in the online edition. Replication files are available in the *JOP* Data Archive on Dataverse (http://thedata.harvard.edu/dvn/dv/jop). Approval for the field experiment was obtained from the Institutional Review Boards of Boston University (protocol 4094X), MIT (protocol 1604551604), and the Universidade Federal de Pernambuco (número de parecer 1571592). Approval for the online survey experiment was obtained from the institutional review boards of Boston University (protocol 4748X) and MIT (protocol 1712167362). The field experiment and the online survey experiment were pre-registered with EGAP (ID 20151118AA and 20180914AA, respectively). The field experiment was part of the Metaketa Initiative on Information and Accountability, funded by Evidence in Governance and Politics (EGAP).

Do voters hold politicians 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, C. Rodriguez, and Artis 2015). Such initiatives assume that information constraints prevent voters from holding elected officials accountable. Therefore, the logic goes, providing relevant and timely information will enable and empower voters to punish bad performers and reward good ones, inducing the selection of better politicians and giving them more incentives to perform.

While the accountability logic is powerful, recent experimental evidence 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, Hidalgo, and Melo 2019b; Dunning, Grossman, Humphreys, Hyde, McIntosh, and Nellis, 2019; Dunning, Grossman, Humphreys, Hyde, McIntosh, Nellis, et al., 2019). Second, some research shows that information prompts accountability voting only for (often small) subgroups of the population or under a specific set of conditions, and not for the electorate in general (Adida et al. 2017, 2020; Bhandari, Larreguy, and Marshall 2020; Boas and Hidalgo 2019). Third, and most normatively worrisome, 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. 2020; Arias et al. 2018; Bhandari, Larreguy, and Marshall 2020; Blattman, Emeriau, and Fiala 2018; Chong et al. 2015; De Kadt and Lieberman 2020). With transparency initiatives increasingly common around the world, it has become urgent to understand what type of information about government performance can foster electoral accountability and for what kinds of voters it can do so.

To address these questions, we study how information about municipal school quality affects voting behavior and electoral outcomes for local incumbents in Brazil, using two complementary research designs. First, we employ a regression discontinuity design to compare municipalities across Brazil that barely met a school quality target and those that barely missed it. Second, we use a field experiment providing information about local school quality to a random sample of voters in the state of Pernambuco. While the quasi-experimental study allows us to measure effects in a naturally occurring environment, thus addressing concerns about general equilibrium impacts, 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 *decrease*, rather than increase, support for the incumbent. 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 line with the pre-analysis plan for the field 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 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. We find evidence consistent with pre-specified hypotheses of voter heterogeneity. Voters who value education react as predicted by political agency models: signals of school quality increase perceptions of spending and improvements in multiple policy areas, including education. However, among voters who give less priority to education, information about school quality has no positive effects, and it appears to *decrease* perceptions of investments and improvements in social assistance. These results suggest that low education voters who more highly value other policy outputs—a majority of the Brazilian population—drive the negative average treatment effects in the natural and field experiments. Our findings are thus consistent with those of Bursztyn (2016), who argues that poor voters in Brazil disapprove of increased educational spending because they prefer cash transfers.

Our results suggest that theories of electoral accountability and retrospective voting, which focus on politicians' competence with respect to valence issues, need to consider the inferences that voters may draw about incumbents' alignment with their own policy priorities. Voters who punish incumbents for "good" policy outcomes may still be seeking to hold them accountable—but for their ability to represent diverse and conflicting interests rather than deliver universal benefits.

INFORMATION, ACCOUNTABILITY, AND UNEXPECTED EFFECTS

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, Grossman, Humphreys, Hyde, McIntosh, and Nellis, 2019). Inspired by this notion, a number of governments and NGOs have developed initiatives seeking to increase citizens' access to information about government performance.

The logic underlying information and accountability initiatives is consistent with the political agency model (Ashworth 2012). In a standard formulation of the model, forward-looking voters seek to make an inference about a politician's "type" based on their observable performance in office (Fearon 1999). A common variant of the model (e.g., Besley 2006) assumes that voters choose based on a single valence issue, which is often portrayed as linked to voter welfare. Voters observe the implementation of policy and make an inference about whether the politician is a "good" or "bad" type. Even the bad type, under certain conditions, can act as a good type and enact policy congruent with voters' desires in order to secure re-election. Where the policy result is imperfectly observable, third-party information generated by the media or auditing institutions can raise voter welfare by strengthening the incentives of bad types to mimic good types, as well as improving voters' capacity to select good types at the ballot box.[1]

An initial body of evidence suggested that information and accountability initiatives can work as predicted by the political agency model (Pande 2011). In Brazil, Ferraz and Finan (2008) show that negative audits of municipal governments reduced incumbents' reelection prospects.[2] 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 predictions of the political agency model. First, some 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, Hidalgo, and Kasahara 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 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, Grossman, Humphreys, Hyde, McIntosh, and Nellis, 2019; Dunning, Grossman, Humphreys, Hyde, McIntosh, Nellis, et al., 2019). Though voters often react strongly to

[1]However, increased information can actually reduce voter welfare in the short term when incumbent re-election incentives are sufficiently weak (Ashworth and Bueno de Mesquita 2014). Informed voters will know the record of the incumbent, which will eliminate the incentive of the bad type to mimic a good type because the incumbent can no longer rely on voter ignorance to let them enact policy closer to their own preferences.

[2]In a follow-up study, however, Avis, Ferraz, and Finan (2018) find that audits discipline politicians mostly through a judicial, rather than an electoral channel.

information on incumbent performance in hypothetical vignettes, they may fail to act on the same information in real life (Boas, Hidalgo, and Melo 2019b; Incerti 2020; Weitz-Shapiro and Winters 2017).

Second, a number of studies demonstrate that delivering information to voters can prompt electoral accountability, but only in particular subgroups or under a unique set of circumstances. 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. 2020). 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 and Hidalgo (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.

Perhaps most troubling for the predictions of the political agency model, several studies have found that information and accountability systems can backfire, with voters punishing incumbents for good outcomes or rewarding them for bad ones. In Benin, informing voters about good performance by their legislator (such as attending and speaking at legislative sessions) prompted punishment because voters assumed a trade-off with particularistic transfers, which they valued more (Adida et al. 2020). 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 another 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 effects. De Kadt and Lieberman (2020) 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, Emeriau, and Fiala (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.

Why might positive information about performance in office prompt voters to punish rather than reward an incumbent? The basic political agency model assumes that voters agree on the desirability of a salient policy outcome. These models also assume that dimensions of performance are positively correlated: information about one valence issue allows voters to infer that politicians are doing well or poorly at managing other issues as well. Yet no issue is inherently a valence or a position issue, and whether voters see a policy outcome as desirable may depend on whether they perceive trade-offs with other policy goals (Stokes 1963, 1992). Incumbents do not have unlimited time, energy, and resources, so they must prioritize their efforts across issues. In this context, good performance in one policy area may have negative implications for other policy areas.

Shifting the focus from unambiguous signals of competence to information about policy priorities suggests rethinking the notion of electoral accountability. Voters who ignore or punish good performance in one policy area are not necessarily behaving irrationally, learning the "wrong" lesson from an information intervention, or failing to hold politicians accountable. Rather, they may value interest representation as much or more than generic competence and reward politicians based on delivery of their preferred policy outcomes (Cruz et al. 2018). Information interventions can generate unexpected effects because voters perceive trade-offs, not only between programmatic and particularistic performance (Adida et al. 2020) but also between different policy priorities.

INSTITUTIONAL SETTING

Brazilian municipal education is a unique setting in which to study electoral responses to public service delivery, since the country has a well run, high visibility system for measuring the quality of public education, one of the most important policy areas for municipal governments. Elections in Brazil's 5,570 municipalities are held every four years in October, and mayors 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 2018. 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 the hands of subnational governments, the federal government plays an important role. In addition to providing funding, it measures education quality through its Basic Education Assessment System (SAEB, *Sistema de Avaliação da Educação Básica*), a set of standardized tests administered across the country. There are two main components to SAEB: the National Literacy Assessment (ANA, *Avaliação Nacional de Alfabetização*), which tests students in third grade, and the National Assessment of School Performance (ANRESC, *Avaliação Nacional do Rendimento Escolar*, also called *Prova Brasil*), which tests students in fifth and ninth grades. ANA is implemented every year, and ANRESC is implemented every two years.

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 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. Targets were defined based on an algorithm that considers baseline levels of performance and are therefore lower for initially weaker schools, municipalities, and states. Once released at the beginning of the period, IDEB targets have not been revised.

By providing an easy-to-understand, binary performance metric (whether targets are met or not), IDEB results are particularly visible and influential. As documented in Appendix A1, 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. As a newer test that does not involve targets, ANA is somewhat less visible than IDEB and ANRESC, though it also attracts media attention after results are released.

There is also evidence of citizen demand for indicators of school performance. As shown in Appendix A2, Google searches for "IDEB" are very common after results are released, even compared to other performance-related terms such as corruption, inflation, and the conditional cash transfer program Bolsa Família. In our survey of voters in the state of Pernambuco, 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, media coverage of the results, and 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 education quality 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 in a naturally occurring environment, without the need for an outside intervention. We expect this binary signal to have strong effects despite the simultaneous release of the underlying continuous scores, given evidence that even highly sophisticated actors in high-information environments, such as financial investors, react strongly to binary signals like credit rating change announcements (Hull, Predescu, and White 2004). Moreover, because the continuous score does not correspond to a meaningful scale, voters and the media would have a difficult time interpreting it without a comparative benchmark.

While numerous factors facilitate electoral accountability for the quality of municipal education services, one issue potentially complicates it: the fact that education is a relatively low-priority area for most Brazilian voters (see Appendix D3). In the AmericasBarometer open-ended survey question about the most serious problem facing the country, education is the least frequently mentioned among the top eight problems, behind security, health, corruption, unemployment, poverty/inequality, the economy in general, and drugs. In our Pernambuco survey, we found similar results when asking about problems in the municipality: school quality ranked sixth on the list of concerns, behind health, crime, jobs, the drought, and sanitation. Nonetheless, education quality remains visible to voters; our Pernambuco respondents reported that it was the second most commonly discussed issue in the 2016 municipal election campaign, behind health.

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 examines the effect of providing information about a municipality's ANA performance on vote for the 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.

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 meeting the IDEB target.

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, we use a continuous measure to increase statistical power and avoid issues with discrete forcing variables in RDDs (Lee and Card 2008). The cutoff is therefore –0.05 in the continuous measure, equivalent to 0 with the rounding applied by the Ministry:

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

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 met the IDEB target) and under control (having missed it). If 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 above and below the threshold:

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

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 might 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.

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. Finally, we exclude observations where separate IDEB results were published for municipal middle schools and municipal primary schools, which could lead to conflicting signals. Appendix A6 presents details of how these and a few other data availability constraints limit our sample.

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:

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

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—unclustered, since the unit of analysis, municipality-period, is also the unit of treatment assignment.

To choose the bandwidth, we use the algorithm proposed by Calonico, Cattaneo, and Titiunik (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, Cattaneo, and Titiunik (2014), which uses kernel weights (putting more weight on observations closer to the cutoff) and corrects for potential bias.[4]

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

[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 use kernel weighting in our baseline specification, "localizing" the regression function using the bandwidth alone, as recommended by Lee and Card (2008, 319).

targets relies on aggregate data, limiting our ability to test mechanisms about how voters process information generated by standardized tests. To complement the RDD, we rely on a field experiment implemented in the state of Pernambuco in partnership with the State Accounts Court (*Tribunal de Contas do Estado de Pernambuco* or TCE-PE), the primary state accountability institution. This experiment, described more fully in Boas, Hidalgo, and Melo (2016, 2019a), provided individuals with information on municipal performance in the ANA 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 2015 IDEB results—necessary to measure change during the mayor's term—were not available until shortly after our study went to the field.

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 mathematics portions of the exam. To compute an overall score, we calculated the mean level of performance for both portions 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 is in Appendix B2. Enumerators also summarized the information orally to maximize information retention and facilitate comprehension among illiterate voters. The flyer design was refined 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.[5] The front of the flyer briefly explained the TCE-PE's auditing responsibilities; the reverse side conveyed municipality-specific details, including a visual illustration of the ranking with comparative metrics. A manipulation check (reported in the last line of the table in Appendix B10) shows that the treatment did improve respondents' knowledge of their municipality's ranking.

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 (Boas, Hidalgo, and Melo 2019a, 2019b). 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

[5]To satisfy legal restrictions on the distribution of campaign advertisements, flyers 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 the Universidade Federal de Pernambuco, but also by lawyers at the TCE-PE. This issue is discussed further in Appendix B3.

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 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.

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

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 (listed in Appendix B5) 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 Appendix B6.

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 (Hainmueller, Mummolo, and Xu 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 Metaketa 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 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." (Boas, Hidalgo, and Melo 2016).

RESULTS

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 (local linear regression with controls), the local average treatment effect of meeting the IDEB target is an 8.5 percentage point decrease in the probability of incumbent re-election ($p < 0.05$), or over 17% of a standard deviation. This result is visualized in Figure 1. The "robust" specification of Calonico, Cattaneo, and Titiunik (2014) returns a similar estimate. Results examining the effect on vote-share (included in Appendix A7) are similar, if noisier due to a lower number of observations.

	Linear	Linear	Robust	Robust
IDEB target met	-0.079*	-0.085**	-0.098*	-0.11**
	(0.045)	(0.043)	(0.053)	(0.056)
Election cycle fixed effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Bandwidth	0.396	0.388	0.396	0.388
Observations	1805	1770	1805	1770

Table 1 – Effect of Meeting the IDEB Target on Re-election of the Mayor

*p<0.1; **p<0.05; ***p<0.01.
Note. The bandwidth is determined by the algorithm of Calonico, Cattaneo, and Titiunik (2014). Standard errors in parentheses are consistent for heteroskedasticity (HC1 in models 1-2, and nearest-neighbor in models 3-4.)

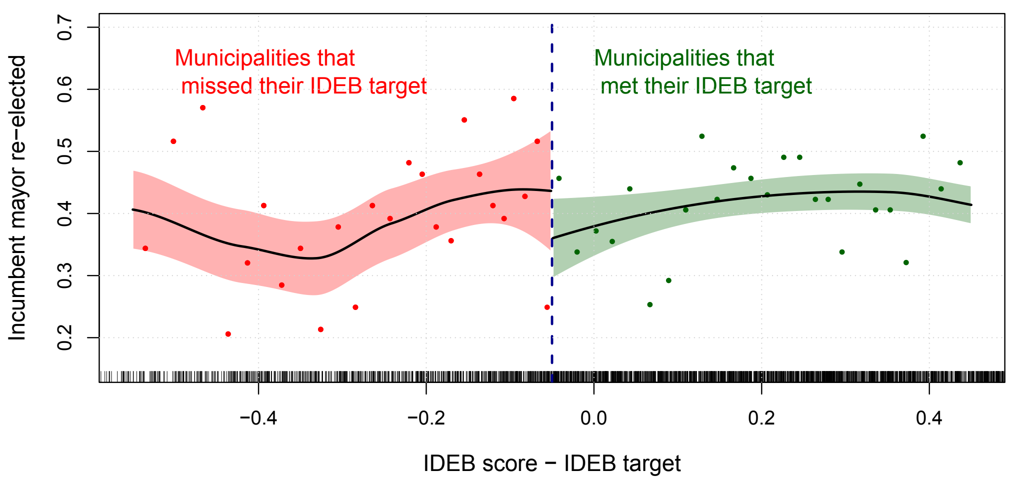

Figure 1 – Effect of Meeting the IDEB Target on Re-election of the Mayor. Dots represent local averages for 50 equally-sized bins. Lines are loess regression lines estimated at both sides of the threshold with no controls. Shaded regions are their 95% confidence intervals.

One concern with RDD results is that they may be dependent on the choice of a particular bandwidth. As shown in Appendix A8, results of our preferred specification have some limited robustness to the choice of alternative bandwidths. While many 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, Cattaneo, and Titiunik (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 RDD shows that meeting the IDEB target has a negative effect on the electoral performance of the mayor—voters appear to punish, rather than reward, improvements in school quality. 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 are stable across specifications. Moreover, the RDD passes placebo tests and there is no evidence that assumptions are violated, lending support to the interpretation of these findings as causal effects.

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 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.

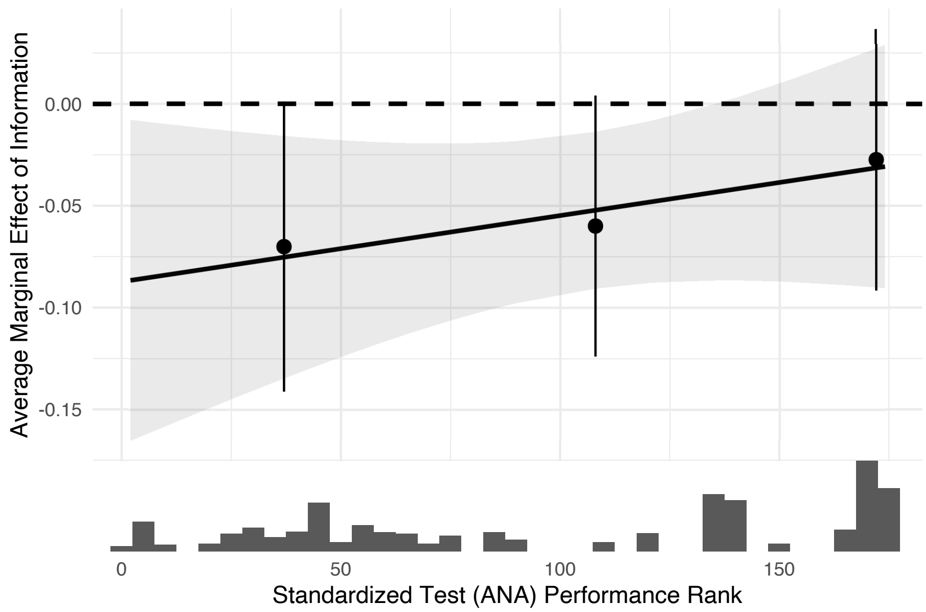

Figure 2 – Effect of Treatment by Educational Performance. Black line is estimated marginal effect of treatment estimated using a linear interaction. Points are effects estimated separately in bins defined by the terciles of ANA Rank. 95% Confidence Intervals are shown. Histogram shows marginal distribution 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 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 subgroup that does have children in local schools, likely due to its smaller size (column 2 in Table 2). However,

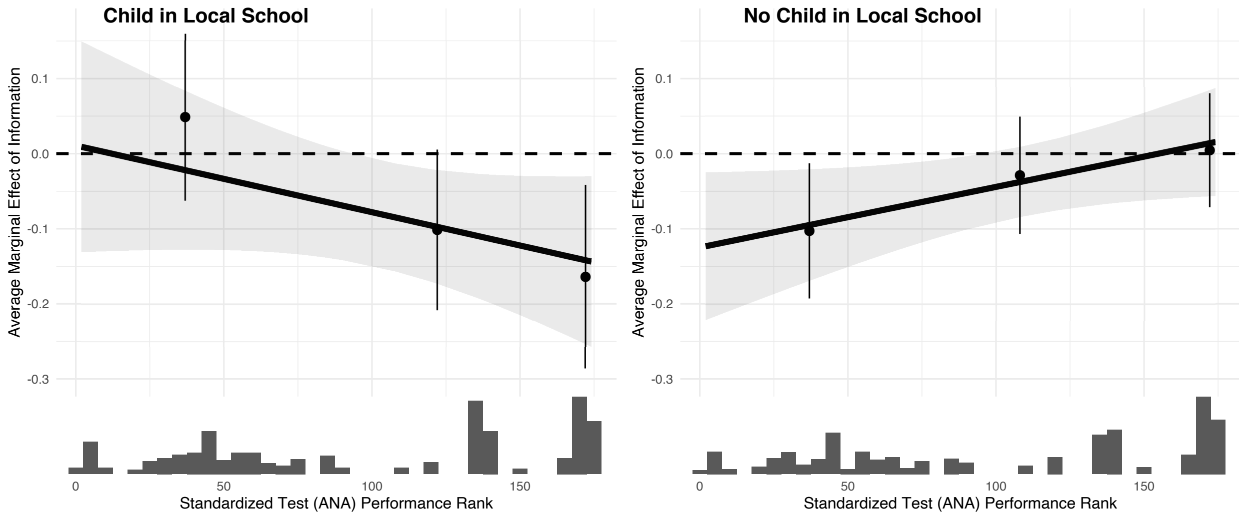

Figure 3 – Effect of Treatment Among Respondents With and Without Children in Local Schools. Black line is estimated marginal effect of treatment estimated using a linear interaction. Points are effects estimated separately in bins defined by the terciles of ANA Rank. 95% Confidence Intervals are shown. Histogram shows marginal distribution of ANA Rank.

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 those who do and do not have children enrolled in municipal schools does not appear to be driven by other correlated observable characteristics. Parents of enrolled children are somewhat less educated, younger, and poorer, but including these variables as additional interactions does not change the relationships observed in Figure 3. Nor does controlling for correlates of municipal characteristics, such as performance on the 2012 ANA exam, alter the results (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. Here, we are also able to document a revealing form of heterogeneity: parents of children enrolled in municipal schools react to information

	All	Parents	Not Parents	All
Treatment	-0.0537^{***}	-0.0713^{**}	-0.0421^{*}	-0.0466^{**}
	(0.0196)	(0.0353)	(0.0239)	(0.0201)
Treatment x Rank	0.0003	-0.0008	0.0008^{*}	0.0008^{*}
	(0.0003)	(0.0006)	(0.0004)	(0.0004)
Treatment x Rank x Parents				-0.0016^{**}
				(0.0007)
Num. obs.	1709	525	1184	1709

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

Table 2 – Effect of Information Treatment.

Note. Covariates omitted. Covariates demeaned so treatment coefficient is estimated average treatment effect. HC2 heteroskedasticity consistent standard errors in parentheses.

about school quality as expected, punishing poor-performing mayors.[6]

TESTING THE MECHANISM

While it is reassuring that parents of children in municipal schools react negatively to deterioration in 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 cash transfers. While there is no empirical relationship between change in educational spending and change in IDEB scores or the likelihood of meeting the target (see Appendix D1), voters may assume such a relationship exists. If so, voters who give less priority to education may punish mayors

[6]Heterogeneity analysis in the RDD suggests that parents' behavior may also have an impact at the macro level. As shown in Appendix A12, the effect of meeting the IDEB target on the re-election of the mayor is reversed in places where there is a larger number of children enrolled in local schools, although this difference is not statistically significant.

who meet their school quality targets because they infer that such mayors are investing less money or effort in more highly valued policy areas. Since most Brazilian voters prioritize issues like security and health more highly than education (see Appendix D3), this dynamic could account for negative average treatment effects of positive information about educational performance.

Research Design

To examine how voters interpret school quality signals, and to test for heterogeneity by priority assigned to education, we draw on an online survey experiment conducted in December 2019. Respondents were recruited via Facebook advertisements (see Appendix C1), a common method for online surveys in comparative politics, especially when estimating treatment effect heterogeneity (Boas, Christenson, and Glick 2020). As summarized in Appendix D2, the online sample was diverse in racial and geographic terms, while being somewhat younger, more female, and much more highly educated than the Brazilian population.

Our treatment informed respondents about whether their self-reported municipality had met its IDEB target when data were released in September 2018. 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 2018, during 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 a photograph of the mayor and a red cross or a green check mark, depending on the IDEB result.

We examine the effect of this treatment on assessments of the incumbent's policy performance and spending choices. The next two batteries of questions asked respondents whether they totally agreed, partially agreed, partially disagreed, or totally disagreed that the mayor had (1) "invested a lot of money in" and (2) "improved the quality of" four different policy areas: education, healthcare, social assistance, and security. Education was always listed first; the order of the remaining items was randomized across respondents but held constant across the two sets of outcome measures.

We estimate treatment effects conditional on whether respondents assign low or high priority to education. Prior to any question about IDEB, respondents were asked to arrange five policy areas—education, health, the economy, social assistance, and security—"according to the priority that you think they should have in the municipal budget." The items were initially presented in random order; the median priority ranking given to education was second place. We score respondents as assigning low priority to education if they ranked it third, fourth, or fifth; 34% of the sample did so.

Estimation and Inference

To analyze the heterogeneous impact of the information treatment on assessments of the mayor, we use the following equation, estimated separately for respondents from municipalities where the IDEB target was and was not met:

$$Y_i = \alpha + \beta T_i + \eta T_i M_i + \pi M_i + \sum_{k=1}^{K} (\theta_k X_i^k + \gamma_k X_i^k T_i + \mu_k X_i^k M_i + \phi_k X_i^k M_i T_i) + \varepsilon_i \qquad (5)$$

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, T_i is an indicator for receiving the treatment, and M_i is an indicator for assigning below-median priority to education. 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 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.

Results

The results of the survey experiment (Figure 4) show that voters who give lower priority to education react differently to signals of improved school quality than those who value education more.[7] For those who prioritize education, positive signals of school quality make respondents more likely to agree that the mayor invested in and improved the quality of all four policy areas (by between 0.18 and 0.29 points, $p < 0.05$). For these respondents, there is support for the predictions of political agency models, as we observe a positive correlation across multiple dimensions of performance.

We find a different pattern among the 34% of respondents who assign low priority to education.[8] Here, positive school quality signals have no statistically significant effect, and they appear to *depress* perceptions of investments in and improvements in social assistance (by –0.08 and –0.11 points, respectively). Though tentative, this finding is largely consistent with the argument by Bursztyn (2016) that poor voters perceive a trade-off between spending on education and on cash transfers.

While the direction of the treatment interaction is consistent with our findings from the field experiment and RDD, the sign of the average and conditional average treatments effects is not. In the survey experiment, those who value education update positively when presented with good news about school quality, and those who do not value education experience only null effects. As a result, average treatment effects on all outcomes are positive in our online sample (Appendix C4). Meanwhile, in the field experiment and RDD, there is no evidence that any group rewards good performance, and many voters

[7] We focus on the treatment effect in municipalities where the school quality target was met. All other pre-specified results are detailed in Appendix C4.

[8] The treatment interaction with priority assigned to education is significant at $p < 0.05$ for all outcomes except healthcare spending.

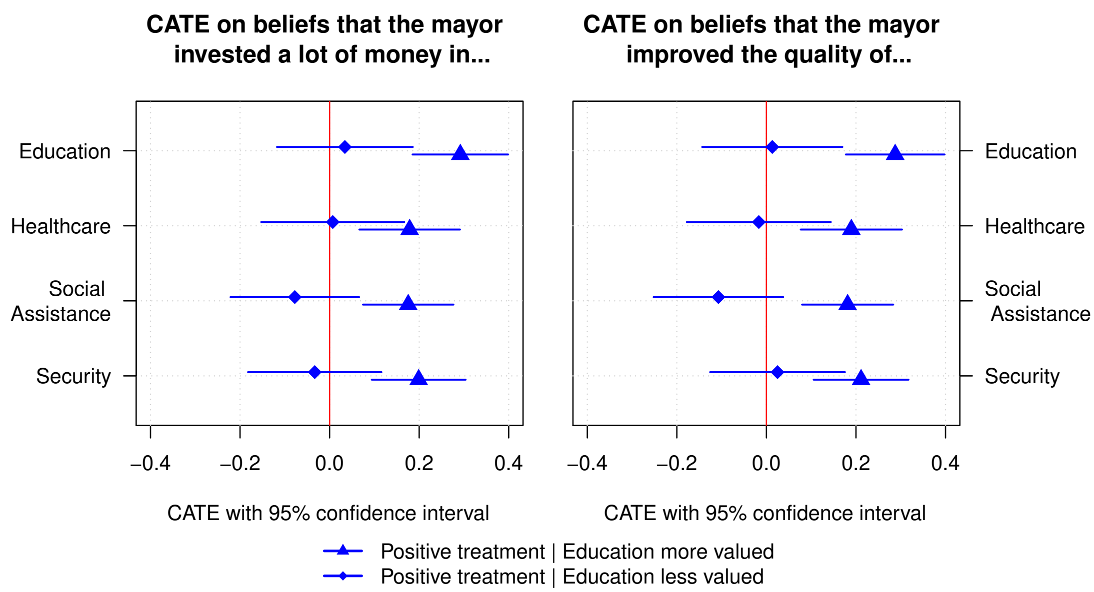

Figure 4 – Conditional Average Treatment Effects (CATE) of receiving positive information about the quality of schools on respondents' agreement with statements about the mayor, by whether they give high or low priority to education. Outcomes are measured in scales that go from 1 ("disagree completely") to 4 ("agree completely"). The "education more valued" group is composed of those who rank education at the median or above it (i.e. among their top two priorities). The "education less valued" group is composed of those who rank at least two policy areas above education.

punish it, adding up to negative average treatment effects.

Several factors might explain the different direction of average treatment effects in the online survey experiment versus the field experiment and RDD. First, our pre-specified approach to defining low priority for education might be too inclusive. When we redefine low priority as ranking this policy area last or next-to-last, we obtain negative point estimates for all conditional average treatment effects, though none are significant given the small size of this subgroup (Appendix C6).

Second, the positive sign of average treatment effects in the online sample might be attributable to the over-representation of highly educated Brazilians who tend to value this policy area. As discussed above and in Appendix D3, few Brazilians consider education to be the most pressing problem facing the country or their municipality, but in our online sample, 26% ranked it as the top priority for municipal spending. Across multiple surveys, a respondent's level of education is the strongest and most consistent predictor of valuing education as a policy area (Appendix D4). As shown in Appendix D2, our online sample is much more highly educated than the Brazilian population, whereas the RDD and field experiment samples are somewhat less educated. If the conditional average treatment effects estimated in the online survey are replicated in the population and these other samples, we should be less likely to obtain positive average treatment effects. Indeed, when re-weighting the online sample to match the distribution of education in the RDD sample, we obtain significant positive ATEs only for investing in and improving the quality of education, not for the other policy outcomes (Appendix C5).

Our findings from the online survey are suggestive rather than conclusive, and future research might seek to uncover stronger evidence regarding the causal mechanism. A new field experiment that administers a real-world informational treatment to a representative sample and measures assessment of mayoral effort and performance in a variety of policy areas would be the ideal research design to test our hypotheses.

With these caveats, our results suggest that the majority of Brazilian voters punish positive educational performance because they infer that incumbents prioritize education over other policy areas that are more important to them. Trade-off thinking may be triggered not only by providing information about policy inputs (Bursztyn 2016) but also by informing voters about policy outputs. 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, Lergetporer, and Woessmann 2018; Busemeyer and Garritzmann 2017).

DISCUSSION AND CONCLUSION

Governmental and non-governmental agencies 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?

Performance-based accountability systems expect that voters will behave according to political agency models, which predict that, after observing performance in office, voters will make an inference as to whether the politician is a "good type" or "bad type" and vote accordingly. Yet cutting-edge research on electoral accountability suggests that the link between information and accountability is weak, with voters often failing to behave as these models predict (Dunning, Grossman, Humphreys, Hyde, McIntosh, and Nellis, 2019; Dunning, Grossman, Humphreys, Hyde, McIntosh, Nellis, et al., 2019). Whether electoral accountability works depends on a number of institutional, socioeconomic and behavioral features, including prior beliefs (Arias et al. 2018), expectations (Gottlieb 2016), socioeconomic endowments (Holbein 2016), coordination (Adida et al. 2020), media markets (Larreguy, Marshall, and Snyder Jr. 2020), and ethnicity (Adida et al. 2017).

In this article, we argue that voters may not act as predicted by political agency models if they seek to hold politicians accountable for their policy priorities in addition to, or instead of, their overall competence. While these models assume that voters agree on the desirability of an area of performance and that different dimensions are positively correlated, in reality voters may perceive trade-offs among issue areas. While education will be a valence issue for some, improvements in education quality are not necessarily good news if they imply less effort or resources being devoted to other policy areas that voters value more. A voter who acts according to inferred policy trade-offs is not necessarily failing to hold politicians accountable; rather, she may seek to reward or punish politicians based on how well they represent her policy interests.

Studying the effect of information about educational performance in Brazil, 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, however, the effect is reversed for parents of children enrolled in municipal schools, for whom school quality should be most salient. Our unique combination of research designs allows us to make inferences about information and accountability at both the macro and individual levels while addressing issues of both internal and external validity as well as potential general equilibrium effects.

Using an online survey experiment to study the mechanism behind these results, we argue that most voters punish politicians for positive educational performance because they perceive trade-offs with other policy areas that they value more. Voters who prioritize the issue of education behave according to the predictions of political agency models, taking positive educational performance as evidence that the politician is a "good type" who also invests resources and provides quality services in other areas. We find no such effect among those who assign low priority to education. Rather, for these voters, positive signals of school quality appear to *decrease* assessments of the mayor's investment in and improvement of social assistance.

These findings have important policy implications. First, policymakers should consider 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 electorate. In some cases, heterogeneity may mean that information prompts performance-based accountability voting only for some small subgroup of the population or under an unusual set of circumstances, but not on a broad enough scale to affect politicians' electoral prospects or induce them to perform better in the future. Second, policymakers and researchers alike should reconsider what it means to hold politicians accountable. Inducing better performance or the selection of more qualified politicians is valuable from a normative standpoint, but so is choosing representatives who have voters' interests in mind and act according to their policy priorities. Accountability interventions that take note of these varied interests and aim to boost the quality of interest representation might stand a better chance of success.

Acknowledgements

Thanks to Marcus André Melo, Marcos Nóbrega, and the State Accounts Court of Pernambuco for their partnership; to Mariana Batista for invaluable help throughout multiple phases of the field experiment; and to Amanda Domingos, Julia Nassar, and Virginia Rocha for research assistance. For feedback on prior versions, we are grateful to Raúl Duarte González, José Ramón Enríquez, Tao Lin, Felipe Nunes, Stuart Russell, seminar participants at MIT, Harvard, Universidade Federal de Pernambuco, and the Southeast Latin American Behavior Conference, and three anonymous reviewers.

References

Adida, Claire, Jessica Gottlieb, Eric Kramon, and Gwyneth McClendon. 2017. "Reducing or Reinforcing In-Group Preferences? An Experiment on Information and Ethnic Voting." *Quarterly Journal of Political Science* 12 (4): 437–77.

———. 2020. "When Does Information Influence Voters? The Joint Importance of Salience and Coordination." *Comparative Political Studies* 53 (6): 851–891.

Arias, Eric, Horacio Larreguy, John Marshall, and Pablo Querubin. 2018. "Priors Rule: When do Malfeasance Revelations Help or Hurt Incumbent Parties." National Bureau of Economic Research Working Paper No. 24888.

Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15:183–201.

Ashworth, Scott, and Ethan Bueno de Mesquita. 2014. "Is Voter Competence Good for Voters? Information, Rationality, and Democratic Performance." *American Political Science Review* 108 (3): 565–587.

Avis, Eric, Claudio Ferraz, and Frederico Finan. 2018. "Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians." *Journal of Political Economy* 126 (5): 1912–1964.

Banerjee, Abhijit, Selvan Kumar, Rohini Pande, and Felix Su. 2011. "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India." Working Paper.

Barber, Michael, Nickolas C. Rodriguez, and Ellyn Artis. 2015. *Deliverology in Practice: How Education Leaders are Improving Student Outcomes.* Thousand Oaks, CA: Corwin.

Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government.* New York: Oxford University Press.

Bhandari, Abhit, Horacio Larreguy, and John Marshall. 2020. "Able and Mostly Willing: An Empirical Anatomy of Information's Effect on Voter-Driven Accountability in Senegal." Working Paper.

Blattman, Christopher, Mathilde Emeriau, and Nathan Fiala. 2018. "Do Anti-Poverty Programs Sway Voters? Experimental Evidence from Uganda." *Review of Economics and Statistics* 100 (5): 891–905.

Bloniarz, Adam, Hanzhong Liu, Cun-Hui Zhang, Jasjeet S. Sekhon, and Bin Yu. 2016. "Lasso Adjustments of Treatment Effect Estimates in Randomized Experiments." *Proceedings of the National Academy of Sciences* 113 (27): 7383–7390.

Boas, Taylor C., Dino P. Christenson, and David M. Glick. 2020. "Recruiting Large Online Samples in the United States and India: Facebook, Mechanical Turk and Qualtrics." *Political Science Research and Methods* 8 (2): 232–250.

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

Boas, Taylor C., F. Daniel Hidalgo, and Marcus André Melo. 2016. "Accountability and Incumbent Performance in the Brazilian Northeast: Pre-Analysis Plan." Pre-Analysis Plan 20151118AA, Evidence in Governance and Politics.

———. 2019a. "Horizontal But Not Vertical: Accountability Institutions and Electoral Sanctioning in Northeast Brazil." In *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I,* edited by Thad Dunning, Guy Grossman, Macartan Humphreys, Susan Hyde, Craig McIntosh, and Gareth Nellis. New York: Cambridge University Press.

———. 2019b. "Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil." *American Journal of Political Science* 63 (2): 385–400.

Buntaine, Mark T., Ryan Jablonski, Daniel L. Nielson, and Paula 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, Leonardo. 2016. "Poverty and the Political Economy of Public Education Spending: Evidence from Brazil." *Journal of the European Economic Association* 14 (5): 1101–1128.

Busemeyer, Marius R., and Julian 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, Marius R., Philipp Lergetporer, and Ludger Woessmann. 2018. "Public opinion and the political economy of educational reforms: A survey." *European Journal of Political Economy* 53:161–185.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. 2019. "Regression Discontinuity Designs Using Covariates." *Review of Economics and Statistics* 3 (101): 1–10.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Chong, Alberto, Ana de la O, Dean Karlan, and Leonard 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.

Cruz, Cesi, Philip Keefer, Julien Labonne, and Francesco Trebbi. 2018. "Making policies matter: Voter responses to campaign promises." Working Paper 24785, National Bureau of Economic Research.

De Figueiredo, Miguel F. P., F. Daniel Hidalgo, and Yuri Kasahara. 2011. "When do Voters Punish Corrupt Politicians? Experimental Evidence from Brazil." Working paper, http://bit.ly/2ByZMqK.

De Kadt, Daniel, and Evan S. Lieberman. 2020. "Nuanced Accountability: Voter Responses to Service Delivery in Southern Africa." *British Journal of Political Science* 50 (1): 185–215.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, Gareth Nellis, Claire L. Adida, Eric Arias, Clara Bicalho, Taylor C. Boas, et al. 2019. "Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials." *Science Advances* 5 (7): eaaw2612.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan Hyde, Craig McIntosh, and Gareth Nellis, eds. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I.* New York: Cambridge University Press.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation,* edited by Adam Przeworski, Susan C. Stokes, and Bernard Manin, 55–97. New York: Cambridge University Press.

Ferraz, Claudio, and Frederico 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, John, and Rosemary McGee. 2013. "The Impact of Transparency and Accountability Initiatives." *Development Policy Review* 31 (s1): s3–s28.

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

Hainmueller, Jens, Jonathan Mummolo, and Yiqing 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, John. 2016. "Left Behind? Citizen Responsiveness to Government Performance Information." *American Political Science Review* 110 (2): 353–368.

Hull, John, Mirela Predescu, and Alan White. 2004. "The relationship between credit default swap spreads, bond yields, and credit rating announcements." *Journal of Banking & Finance* 28 (11): 2789–2811.

Humphreys, Macartan, and Jeremy Weinstein. 2012. "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda: Preliminary Analysis." Working paper, http://bit.ly/2kmrljT.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–635.

Incerti, Trevor. 2020. "Corruption Information and Vote Share: A Meta-Analysis and Lessons for Experimental Design." *American Political Science Review.*

Larreguy, Horacio, John Marshall, and James M. Snyder Jr. 2020. "Publicizing malfeasance: When the local media structure facilitates electoral accountability in Mexico." *The Economic Journal.*

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

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

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

Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57 (2): 368–377.

———. 1992. "Valence Politics." In *Electoral Politics,* edited by Dennis Kavanagh. New York: Oxford University Press.

Weitz-Shapiro, Rebecca, and Matthew 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.

Taylor C. Boas is Associate Professor, Department of Political Science, Boston University, 232 Bay State Road, Boston, MA 02215 (tboas@bu.edu). F. Daniel Hidalgo is Cecil and Ida Green Associate Professor of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E53-470, Cambridge, MA 02139 (dhidalgo@mit.edu). Guillermo Toral is PhD candidate in Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E53-470, Cambridge, MA 02139 (gtoral@mit.edu).