

Left Behind? Citizen Responsiveness to Government Performance Information

JOHN HOLBEIN *Duke University*

Do citizens respond to policy-based information signals about government performance? Using multiple big datasets—which link for the first time large-scale school administrative records and individual validated voting behavior—I show that citizens react to exogenous school failure signals provided by No Child Left Behind. These signals cause a noticeable increase in turnout in local school board elections and increase the competitiveness of these races. Additionally, I present evidence that school failure signals cause citizens to vote with their feet by exiting failing schools, suggesting that exit plays an underexplored role in democratic accountability. However, performance signals elicit a response unequally, with failure primarily mobilizing high propensity citizens and encouraging exit among those who are white, affluent, and more likely to vote. Hence, while performance signals spur a response, they do so only for a select few, leaving many others behind.

We are transforming our schools. . . . We are insisting on accountability, empowering parents. . . and making sure that local people are in charge. We will leave no child behind.

President George W. Bush¹

Information matters in nurturing democratic accountability. When citizens are informed, they can play an active role in holding government accountable; when they are not informed, they are less likely to act when government performance deteriorates—resulting in policy that drifts further from public preferences (e.g., Schlozman, Verba, Brady 2012). Information, in short, “enables [citizens] to impose sanctions on . . . power-wielders” and, as such, enhances accountability (Grant and Keohane 2005, 30). Unfortunately, the citizenries of many contemporary democracies have strikingly low and unequal levels of information. While this problem it is widely understood (e.g., Delli Carpini and Keeter 1996; Ferejohn 1986; Przeworski, Stokes, and Manin 1999), it remains unclear how best to address this issue.

Increasingly common are policy-based attempts to address low and unequal levels of citizen information through *performance accountability systems*. Performance accountability systems differ in their form and substance, but generally have three components: measurement, sanctions, and publication. In the first, government performance is placed on a common standard (e.g., letter grades from A to F). In the second, those who do not meet these standards are punished. In many cases, the main punishment comes from the third component—distribution of performance to the public. These performance accountability mechanisms are intended to provide citizens with clear metrics of policy performance to enable responsiveness. In

recent years, performance accountability systems have become increasingly common across a variety of sectors—domestic and international, public and private. For example, in the United States, some governments have begun to measure and publish performance grades for municipal agencies.² These types of reforms have also spread widely through international development (e.g., politician audits), financial (e.g., disclosure requirements in banking), food services (e.g., letter grades of cleanliness posted on restaurants), child-care (e.g., quality rating and improvement systems), transportation (e.g., A+B performance-based contracts), environment (e.g., performance-based air quality standards), and education contexts (Björkman and Svensson 2009; Stecher et al. 2010). The No Child Left Behind Act of 2001 (NCLB) is a prominent example of a public policy designed with a performance accountability structure. This seminal law provides citizens clear signals about which public schools are “failing” that are designed to help citizens hold local school officials accountable, as the opening quote from President Bush makes clear.

Despite the widespread trend towards performance accountability policies, exemplified by NCLB, there lacks a comprehensive empirical examination of citizen responsiveness to performance information signals. Do citizens respond to NCLB-based signals information that their local school is failing? The answer to this question is theoretically unclear as it depends on citizens’ underlying levels of information—if residents are fully informed we might not expect to see a response, but if they instead have limited information, failure could serve as an alarm that something is amiss. Similarly, if citizens do respond, how do they do so: with increased voice in their local school board elections or by exiting the failing school? Finally, who is most likely to respond? In other words, does performance information serve to narrow or exacerbate inequalities in the electorate?

In addressing these questions, I bring together three unique “big data” sources from North Carolina. These combine detailed panel information on 15 million

John Holbein is Ph.D. Candidate, Sanford School of Public Policy, Duke University (john.holbein@duke.edu).

I wish to thank the National Science Foundation (#SES-1416816) for their generous financial support. I would also like to give special thanks to David Campbell, Nicholas Carnes, William Darity, James Hamilton, Andrew Heiss, Sunshine Hillygus, Helen Ladd, Douglas Reed, Ken Rogerson, Danielle Vance, and Jacob Vigdor for their invaluable comments.

¹ Quote taken from the 2004 Republican Convention (Sept. 2, 2004).

² In 2014, New York City began grading agencies on their fiscal prudence (see NYC PR 14-09-78).

citizen-year observations from the state's voter file, 5 million student-year observations in the state's public school system, and the performance of the state's 15,000 school-year observations. With these data, I examine the causal impact of school failure signals on metrics of citizen accountability, including voter turnout in school board elections, competitiveness of school board elections, and exit from failing schools. I do so with a regression discontinuity design (RDD) that leverages exogenous variation around the arbitrary school failure cutoff. Enhancing this RDD, I leverage the panel nature of my data to make an even stronger comparison: between marginally failing schools and the same schools when they marginally pass.

This analysis shows that citizens respond to school failure signals, and noticeably so. After a school is labeled as failing by NCLB, voter turnout in subsequent school board elections rises substantially—by five to eight percentage points on average. In addition, I find that failure signals increase the competitiveness of local board elections. In short, it appears that performance signals prompt residents to use their voice in an attempt to hold local elected school officials accountable. Beyond this voice response, I also show that—as theory has long predicted, but few empirical studies have corroborated—citizens respond by voting with their feet, exiting communities that experience school failure. Such responses demonstrate that citizens may not always be attentive to government performance, but instead react when performance crosses a performance threshold and a proverbial alarm sounds. Despite this response, these voice- and exit-based responses vary in important ways. These gains come largely from those most likely to participate at baseline based on previous vote history, socio-economic status (SES), and race. Thus, while performance-based accountability encourages citizen responsiveness, it does so unequally—promoting the strongest response among those already engaged.

This examination contributes to a salient policy debate as policymakers increasingly look to performance-based systems as a means of increasing transparency, promoting local accountability, and ensuring “good governance” (Lewis-Faupel et al. 2014; Pande 2011). By using data from NCLB, this study provides direct information about the causal impact of a current, large-scale public policy initiative on citizen action. Further, this article shows that accountability studies should expand to consider not just metrics of voice but also of exit, which constitutes an important, yet underexplored, venue of citizen responsiveness. Finally, these results highlight that performance accountability policies can be a double-edged sword. Although they can promote citizen responsiveness, they may also exacerbate participatory inequalities.

BACKGROUND & THEORETICAL FRAMEWORK

In recent years performance accountability systems have spread widely across health, transportation, agri-

culture, penal, environment, international development, and education sectors. These systems provide citizens with information signals about government performance. Although designed to improve democratic accountability, not much is known about if and how people respond to these performance signals. Indeed, one of the longest standing debates in the field is about the basic capacity of citizens to use information in their political decision-making. Given that, will performance information prompt citizen action or will the “sheer access to current information,” ensure that “the great bulk of . . . information dispensed . . . blows away like chaff” (Campbell et al. 1980, 254)? The existing literature on this point has been decidedly mixed.

Previous research suggests that voters are sensitive to economic performance (e.g., Fiorina 1978; Rudolph 2003), crime (e.g., Arnold and Carnes 2012), military deaths (e.g., Grose and Oppenheimer 2007), and distributive spending (Chen 2013), but there is little consistent evidence that citizens rely on performance metrics from specific policies. For example, some find that citizens respond to performance information (e.g., Banerjee et al. 2010), but others find the opposite (e.g., Olken 2007).³ Research on school performance, in particular, shows conflicting results in this regard. Using data from South Carolina, Berry and Howell (2007) find that poor public school performance predicts metrics of citizen responsiveness in school board elections. In contrast, using observational data from multiple states Rhodes (2014) finds that more stringent performance reporting requirements predicts lower levels of citizen responsiveness.

These conflicting results may be because many studies linking government performance and citizen responsiveness struggle, to varying degrees, with endogeneity. Unfortunately, exogenous changes in government performance are few and far between. In the case of public school performance, this is especially problematic because geographic proximity to a high or low performing public school is not allocated randomly. As a result, the relationship between metrics of performance and citizen responsiveness may be subject to omitted variable bias or selection effects (Ashworth 2012). As such, gaining leverage on the question of whether citizens respond to performance information requires a robust causal identification strategy.

In this article, I leverage a natural experiment with No Child Left Behind (NCLB) that avoids these endogeneity issues. Under this prominent performance accountability reform, local public schools are labeled “failing” when they do not meet an arbitrary performance cutoff (I outline the nature of this cutoff in detail below). While schools do not fail randomly, a group of schools near the failure cutoff appear to cross over it—or not—in an as-good-as random manner (Ahn & Vigdor 2014a; Holbein and Ladd 2015). Moreover, as NCLB has been in place for some time, schools fall

³ For an overview of the mixed findings from international settings, see Pande (2011).

on either side of this arbitrary cutoff repeatedly, allowing even stronger comparisons across the failure cutoff within individual schools. This setup allows for a compelling examination of whether performance information elicits a response, in this case by increasing citizens' turnout in and the competitiveness of local school board elections.

In addition to allowing an exploration of whether citizens respond, NCLB provides an opportunity to reconsider *how* citizens might respond to information about government performance. Previous accountability studies have predominantly focused on one type of citizen response—citizens' use of voice (typically their vote) to alter who is in power. However, this focus may be overly narrow. Theoretical models have long predicted that when government performance deteriorates citizens will use both voice and exit (Hirschman 1970). Conceptually, scholars have identified exit as being strongly linked to democratic accountability. The thought is that if citizens move from a municipality when economic performance declines, for example, public agencies will lose revenues because of a diminished tax base. In such cases, elected officials might have to cut services, which can then snowball into further performance declines and more citizen dissatisfaction (Hirschman 1970, 21–30). Moreover, exit may be damaging to the reputations of elected officials, acting as a signal of their own low performance. Such an outcome enhances the likelihood that the elected official is voted out of office (Warren 2011, 694). Hence, exit, or the threat of exit, is thought to incentivize high performance from elected officials (Chubb and Moe 1991; Friedman 1955).

For these reasons, some have argued “exit-based empowerments should be as central to the design and integrity of democracy as distributions of votes and voice” (Warren 2011, 683). However, while many theoretical models consider both voice and exit as important to ensuring “representative outcomes in local politics,” few empirical studies consider both behaviors in tandem (Dowding and John 2008; Trounstein 2010, 408).⁴ This is unfortunate due to the presence of numerous venues where exit can be used. For example, if citizens are dissatisfied with the government benefits they are receiving, they may exit to a community with better benefits (Tiebout 1956). Alternatively, citizens may exit if they have a strong aversion to a certain policy, party, or politician.⁵ More broadly, whenever a choice set and competition exists between public and private providers—as it does in health, transportation, and housing sectors, to name a few—people can choose to exit as a signal of their dissatisfaction. In the education context, this option is particularly salient: being an explicit option under recent school choice reforms,

⁴ A few works come close. Henderson (2010) considers the impact of school performance on exit patterns, but not voice, while Cox and Witko (2008) examine the relationship between exit and voice.

⁵ As some have done in response to tax systems (see “Why More Americans Are Renouncing U.S. Citizenship,” NPR, 2/20/2014) or threatened to do with the election of Presidents Bush and Obama. (see “‘I’m Moving To Canada,’ And Other Recent Overreactions To Political Events,” Medialite, 9/11/2012).

like school vouchers and charter schools. Moreover, it plays a central role in recent performance accountability systems, like the school transfers allowed under NCLB. Given its potential as a response option, studies that ignore exit may produce misleading results. If, for example, performance signals promote exit but not voice, studies that examine voice-based responses in isolation will come to the wrong conclusions.⁶

As exit has long been ignored empirically, many uncertainties remain as to the nature of this type of citizen response. For example, previous research is ambiguous as to whether exit is a tool for those marginalized or those already likely to participate. On the one hand, some theories predict that exit will “atrophy the development of the art of voice” by providing a substitute for high propensity participators (Hirschman 1970, 43; see also Rich and Jennings *forthcoming*). However, others predict that “exit can function as [a] low-cost, effective empowerment, particularly for those without voice” (Warren 2011, 683). Exploring who exits, in short, is important as it may have implications for citizen voice and the responses of elected officials.⁷

Finally, the question of who exits strikes at a broader ambiguity from previous accountability studies. For the most part, this literature has very little to say regarding *who* responds to information about government performance. Most consider performance information, if anything, to have a uniform impact. However, work from other studies of voter behavior hints that such a model may be an oversimplification (e.g., Arceneaux and Nickerson 2009; Enos, Fowler, and Vavreck 2014). For example, some work suggests that information may spur the involvement of those least likely to participate (e.g., Di Gennaro and Dutton 2006). This may occur if information has diminishing returns (e.g., Lassen 2005; Wolfinger, Highton, and Mullin 2005). Under this response heterogeneity, providing performance signals—school failure signals, in the case of NCLB—would reduce inequalities in citizens' use of voice and exit. Conversely, however, there are reasons to suspect that performance signals may perpetuate inequality in these outcomes. This may occur if recognition and responsiveness are contingent on individuals' ability to process new information. Indeed, this explanation is central to the idea of a knowledge gap from the political campaigns literature, which predicts that new information technology will enhance participatory inequality (Holbrook 2002). Given the current state of the accountability literature, however, which

⁶ Another reason to examine exit and turnout involves simple math. As turnout in an area equals $\frac{\text{votes}}{\text{population}}$, changes in turnout can occur when population shifts, even when the number of votes does not.

⁷ Exit heterogeneity could damage or enhance to accountability. If high propensity individuals use exit, elected officials could be more likely to respond, because they lose the social and financial contributions of these residents. Alternatively, elected officials could be less likely to respond because the particularly active people have exited the area, and, as a result, the chances of being removed from office diminish. This ambiguity deserves direct attention. Doing so here is simply beyond the scope of this article. I seek to establish whether exit occurs; another valuable question is how exit affects elected officials.

of these scenarios is realized remains theoretically indeterminate.⁸

In sum, there remains much to be understood about the effectiveness of performance accountability policies, including if citizens respond to performance information, how they respond, and who is most likely to do so. Relying on a diverse set of administrative data and a compelling statistical approach, I use No Child Left Behind to examine these important topics.

EMPIRICAL CASE: NO CHILD LEFT BEHIND

Signed into law on January 8, 2002 as a bipartisan reform, No Child Left Behind is widely considered the “most far-reaching education policy . . . over the last four decades” (Dee and Jacob 2011, 149). Since its implementation, NCLB has been the primary means for improving education outcomes. By many accounts, the law has fundamentally altered how public schools in the United States operate (Dee and Jacob 2011; Holbein and Ladd 2015; McDonnell 2009). The law has been a subject of intense debate, with members of Congress almost 15 years later continuing to argue over its merits.⁹

No Child Left Behind is explicitly focused on raising overall performance and, as the law’s name implies, reducing inequality. To accomplish these goals, the law mandates that schools implement performance-based accountability systems. Under NCLB’s performance system, measurement consists of administering standardized tests to students. Publishing performance occurs through the labeling schools as “failing” if arbitrary proficiency thresholds are not met. Dissemination of this information is achieved through letters from local school officials, an official website that receives substantial web traffic, and various informal channels (see Section I in the Online Appendix). Nearby residents, regardless of whether they have children in schools, have reason to pay attention to these failure signals, as these play an important role in influencing housing values (Figlio and Lucas 2004).

Research on the impact of NCLB has focused primarily on its role in influencing student test scores, finding mixed results (Holbein and Ladd 2015). However, beyond promoting student achievement, NCLB has another less-examined purpose. Following standard accountability theory, NCLB was designed, in part, as a mechanism for increasing local accountability. When designed, it was hoped that this reform would empower communities to hold local elected school officials accountable (McDonnell 2009). As evidence of

this, the text of NCLB repeatedly mentions its intention to “lower barriers to . . . participation” as a means of putting pressure on local officials (e.g., ESEA 2002, 115 STAT. 1456). The law’s designers argued that providing performance information would serve to put local “school boards . . . on notice.”¹⁰

Despite this objective, little published work has examined the impact of NCLB on measures of democratic accountability. The work that has been done has focused on citizen attitudes rather than behavior; for example, Rhodes (2014) explored some of the implications of NCLB for citizen efficacy, while Chingos, Henderson, and West (2012) explored failure’s impacts on subjective evaluations of schools. None have linked school performance and validated individual behavior. Further, no study has explored how citizens may respond—through voice, exit, or both—and who is likely to respond.

DATA

To explore these topics, I use a combination of unique “big data” sources. The approach used here links—for the first time—a rich set of administrative data from public schools and election administration records. These observations come from a single state: North Carolina. This is done due to the richness of student, school, and registered voter data in the state. Unlike many other states, North Carolina has long collected student-level information for all students in public schools, having almost 20 years worth of individual-level student data. As described here, I use data from the state that combines a large population of public schools over time ($\approx 15,000$) and student-year observations ($\approx 5,000,000$) with the validated voting behavior of a large sample of registered citizen-year observations ($\approx 15,000,000$). Bridging these large datasets, I create a unique set of information that can be used to document—among other things—the impact of school failure signals on metrics of democratic accountability such as voter turnout in school board elections, competitiveness in those elections, and exit from public schools.¹¹

In this analysis, the independent variable of interest is whether or not schools failed to make adequate yearly progress (AYP) under NCLB.¹² The failure determination is described in detail in the methods section below, as it has direct bearing on the identification strategy used. Generally speaking, AYP failure is determined by low student performance on standardized tests. Low performing schools signal to their surrounding communities that they have failed, while higher performing schools do not.

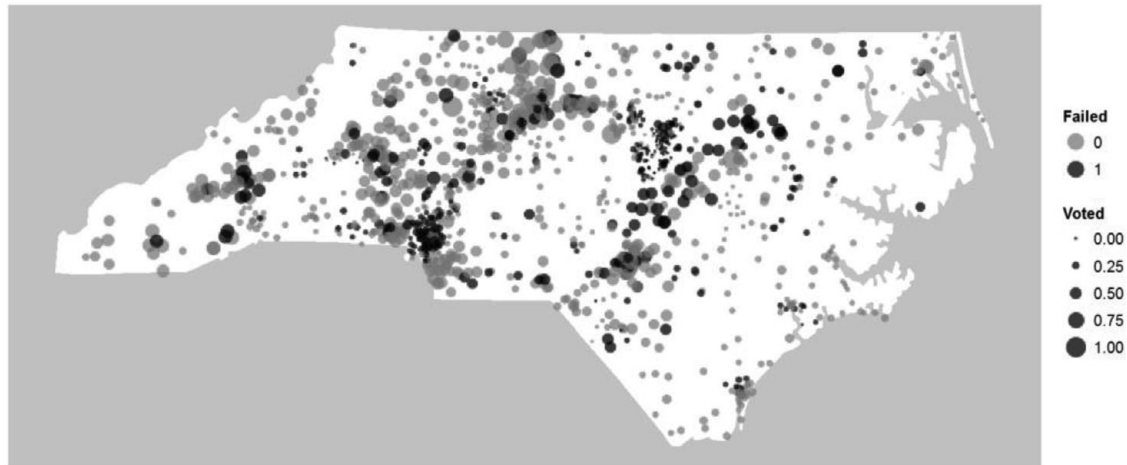
⁸ Another heterogeneity of interest is along latent dimensions of support for those in power. Those who support the incumbent may be demobilized by failure, while those who do not may be mobilized. Unfortunately, I cannot explore this in my local context, due to the nonpartisan nature of the races I explore and the lack of available data on incumbency.

⁹ The performance-based system No Child Left Behind created remains in force as of July, 2015. President Barack Obama granted waivers to several states in 2012, but required that these keep some form of performance accountability. Various revision bills have been brought to the floor of Congress as recently as July 2013 and July 2015.

¹⁰ Rep. George Miller, Congressional Record of the 107th Congress (2001–2002), p. H2594.

¹¹ Focusing on a single state has drawbacks. However, using North Carolina mitigates some of these. As the tenth most populous state in the United States, the state has a diverse population living in both rural and urban settings and voter turnout that is comparable to overall national rates.

¹² This is drawn from North Carolina school performance data from 2003–2011.

FIGURE 1. Performance Failure and Voter Turnout

Notes: The map displays the performance of elementary schools in North Carolina in the 2009–2010 school years matched to turnout information in the next local election.

To examine how school failure signals influence metrics of citizen responsiveness relevant to democratic accountability, I use four outcomes. The first measure—voter turnout—is often used in accountability studies (e.g., Banerjee et al. 2010; Chen 2013; Chong et al. 2011; Pande 2011). In this study, I restrict the turnout measure to elections where a school board race is on the ballot, when local school performance is most salient to the vote at hand. In North Carolina, these elections generally occur during May, even-year primaries.¹³ In addition to turnout, I examine the competitiveness of school board elections, including the number of candidates running and the margin of victory.¹⁴ These are also commonly used in accountability studies (e.g., Berry and Howell 2007; Ferraz and Finan 2008; Niemi et al. 1995), particularly in local races where data on incumbency is often not available.¹⁵ Finally, I examine exit, or voting with one’s feet, which has rarely been used in accountability studies. This outcome is measured using the school enrollments and exit data housed at the North Carolina Education Resource Data Center (NCERDC). This large-scale administrative dataset documents the flow of students into and out of the North Carolina public schools, and is readily matchable to public school performance.¹⁶

The matching of school performance data to voter data requires some work. These are collected at different levels—the unit of observation in the voter file is the individual, while the unit of observation in the accountability data is the school. Unfortunately, voter

files generally do not indicate specific school assignment and official school boundary maps are limited in their availability and quality. Thus, to fit the two data sources together, I matched citizens to the school that minimized the Euclidean distance (as the bird flies) between home addresses and public schools.¹⁷ This large-scale matching process was done four times: identifying an elementary, middle, and high school for each voter, and one closest among the three.

This approach comes with distinct advantages. It allows for a breadth that previous studies linking schools and citizens have not possessed. Moreover, robustness checks indicate that geographic matching approximates assignment matching sufficiently for the identification strategy used. Importantly, geographic matching does equally well on either side at the school failure cutoff. Put differently, those schools that marginally fail are no more likely to be matched to an assigned school than marginally passing schools. This makes it very unlikely that the geographic matching procedure biases the regression discontinuity estimates outlined below. In short, the matching procedure likely only introduces additional noise into these estimates: making it harder to find an effect, if one is indeed present.¹⁸

Figure 1 illustrates the unique dataset this matching procedure produces. It maps school performance in 2009–2010—a representative year—and voter turnout in the next school board election. In the map, the points represent individual public schools. These are shaded by performance under NCLB and sized by turnout in the school zone. From the figure it appears that failing

¹³ I use turnout data from elections held 2004–2012.

¹⁴ In North Carolina, school board outcomes are available primarily from 2008 to 2011.

¹⁵ Unfortunately, incumbency is not available in the North Carolina school board outcomes data.

¹⁶ I obtained these administrative files through the NCERDC’s standard data application procedure.

¹⁷ Geographic matching approach is becoming increasingly common. Several works including Barrows (2014), Minkoff (2014), and Kogan, Lavertu, and Peskowitz (forthcoming) utilize this technique.

¹⁸ The specification checks outlined here are discussed in the Online Appendix (see Section IV).

schools actually *decrease* turnout in school board elections. This can be seen by the preponderance of large grey dots (passing schools with high turnout) and relative lack of large black dots (failing schools with high turnout). Bivariate models confirm this simple eyeball test: showing that schools which fail have voter turnout that is 8% lower than schools that pass.¹⁹

This difference is intuitive, but not terribly informative of performance information's effect. Even if we were to include a host of potentially important controls, schools that fail would likely be different than schools that pass on a number of unobservable dimensions. Perhaps most importantly, the relationship between performance signals and voter turnout could be endogenous—voter turnout may be strongly related to parent involvement or social capital, both of which are thought to influence school performance (e.g., Helliwell and Putnam 2007; Jeynes 2007). In short, this correlation between the performance and citizen response may be highly misleading. Getting around this problem and isolating performance information's effect requires some form of causal identification strategy.

METHODS

To estimate the unbiased impact of performance signals on citizen voice and exit, I leverage a discontinuity in school performance at NCLB's school failure cutoff. As has been well established, observations that are sufficiently close to an arbitrary discontinuity are separated primarily by exogenous shocks (e.g., Butler and Butler 2006). Regression discontinuity models leverage this exogenous variation, using data on either side of a discontinuity to establish treatment and control groups that are similar on observables and unobservables. Under modest assumptions, RDD models produce unbiased local average treatment effects (e.g., Lee and Lemieux 2010).²⁰

Regression discontinuity models require two parameters: first, treatment, and second, the running variable. In the application used here, treatment consists of a public school failing to make adequate yearly progress (AYP), sending a signal to the surrounding community that the school has been labeled failing. Control schools—those that marginally make AYP—receive no such signal. Under NCLB, AYP status is determined by the proportion of students who score at proficiency on standardized tests. The basic idea behind determining AYP status is that when too few students reach proficiency, the school fails. NCLB complicates this slightly by requiring that all student subgroups reach specified cutoffs; if one subgroup fails, the entire school fails.²¹ Provisions allowing exemptions for passing further complicate the determination

failure. These exemptions include passing with growth (improving sufficiently from one year to the next), or passing with interval (being sufficiently close to passing). If either of these exemptions pulls a subgroup above the failure cutoff, that subgroup passes. Despite this complicated formula, determining school failure is straightforward given that school performance is made public—we know very clearly the schools that failed and those that passed.

In contrast, the formula determining AYP status makes specifying the running variable—or how close a school is to failing—difficult (but not impossible). Traditional regression discontinuity approaches use a single metric for the running variable. In the NCLB application, with 20 subgroup scores and multiple channels for passing with each subgroup, the rule used for choosing the running variable must account for both of these features. To do so, I use the procedure developed by Ahn and Vigdor (2014a), which mirrors other approaches to specifying the running variable with multiple inputs (e.g., Jacob and Lefgren 2004; Matsudaira 2008).²² This approach follows the intuition behind the codified rules of NCLB. The first step chooses one channel (overall, growth, or interval) for each subgroup in the school. The decision rule used here is to choose the channel that gives each subgroup the highest score. The intuition behind this decision rule is that, under NCLB's rules, if any channel places the subgroup above the threshold, that subgroup passes. Thus, the channel that produces the highest score identifies how far a school's performance would have to deteriorate to not pass on a given subgroup.²³ In the second step, one subgroup score is chosen to represent how close the school was to failing. The decision rule for the subgroup score is to choose the minimum subgroup score. The intuition behind this decision rule is that, under NCLB's rules, if any subgroup score falls below the cutoff, the school fails. If schools are failing, passing occurs only once all subgroup categories are brought above the threshold. Thus, the minimum subgroup score approximates how far a failing school has to improve to pass. (An example following these two steps for an individual school is available in the Online Appendix.)

This approach to specifying the running variable is more accurate than other naïve approaches. With the approach just outlined, I am able to correctly predict $\approx 80\%$ of schools as either passing or failing; whereas, if I simply average across the subgroups I am able to correctly predict only $\approx 50\%$ of schools' performance.²⁴ Misidentification of school failure status does sometimes occur: in a handful of schools, the running variable indicates that a school failed when we know from

¹⁹ Based on a bivariate model of turnout on proximity to failure among the closest schools match.

²⁰ I explore some of these assumptions in the Online Appendix. In short, these checks generally suggest that the school failure discontinuity is a valid source of exogenous performance information.

²¹ The 10 subgroups include All students, American Indian, Asian, Black, Hispanic, Two or More Races, White, Economically Disadvan-

taged, Limited English Proficiency, and Students with Disabilities. These groups are measured in both reading and math, for a total of 20 subgroup scores.

²² Some have begun to grapple with using multiple measures determining multiple treatments (Papay, Willett, and Murnane 2011). However, I use multiple measures to determine a single treatment.

²³ This approach assumes that all channels improve (or deteriorate) in an order-preserving fashion.

²⁴ Based on the first-stage estimates from the RDD models (see Online Appendix, Section VIII).

the public data that the school was marked passing, or vice versa. Such misidentification comes primarily because of ambiguity in the interval exemption and in the other auxiliary measures used to determine school failure status.²⁵ Because I am not able to perfectly categorize schools with the proximity measure, fuzzy RD is required. This approach is standard in applications with multiple inputs determining the running variable (e.g., Ahn and Vigdor 2014a; Matsudaira 2008).

Fuzzy RDDs use an instrumental variables approach to adjust for noncompliance (Angrist and Pischke 2008; Hahn, Todd, and Van der Klaauw 2001; Imbens and Lemieux 2008). Like in randomized-control experiments, Fuzzy RDD corrects for noncompliance by using treatment assignment as an instrument for treatment receipt. In the NCLB application, treatment assignment is predicted failure (based on the running variable) and treatment receipt is whether or not a school actually failed to make AYP. The limited noncompliance in this application comes from schools that are, as best we can tell, marked failing when they should actually be passing, or vice versa. Still, as noncompliance of this type is relatively rare, the instrument is sufficiently strong to satisfy the assumptions of IV models (Stock and Yogo 2005).²⁶

Equations (1) and (2) show a simplified form of my fuzzy RD models. Each of the variables in the model is indexed at the individual (*i*), school (*s*), or year (*t*) level.

$$F_{st} = \gamma_0 + \gamma_1(P_{st}) + g(R_{st}) + \gamma_3(X_{ist}) + \delta_s + \xi_{ist}, \quad (1)$$

$$Y_{it} = \beta_0 + \beta_1(\hat{F}_{st}) + g(R_{st}) + \beta_3(X_{ist}) + \delta_s + \varepsilon_{ist}. \quad (2)$$

In the first stage, actual AYP failure status (F_{st}) is estimated as a function of the running variable $g(R_{st})$ —which I model using a quartic polynomial, but with other parametric and flexible nonparametric specifications in robustness checks—and the excluded instrument determined by the running variable (P_{st}). The simultaneously estimated second stage produces the causal effect of signaled failure on the outcomes of interest (Y_{it}): school board turnout, electoral competitiveness in board elections, and school exit.²⁷

Also included in the models are a vector of time varying covariates (X_{ist}) and a school fixed effect (δ_s). Including a school fixed effect is similar to combining a regression discontinuity with a difference-in-difference approach (e.g., Holbein and Hillygus, forthcoming;

Jacob and Lefgren 2004). This approach leverages variation within schools' performance across the running variable over time. This protects against the limited possibility that schools on either side of the failure cutoff are different in unobserved ways. While specification checks provided in the Online Appendix indicate this is likely a minimal problem, this is done out of an abundance of caution for potential unobserved imbalances unique to the failure cutoff. In addition to increasing internal validity, this approach allows us to estimate a compelling change. It compares how citizens react when their school goes, for example, from nearly failing in one year to actually marginally failing in another. With the RDD and the school fixed effect, this approach holds constant unobserved factors that might bias the relationship of interest.

Despite the causal leverage this RDD design allows, the focus on a discrete failure signal might make it more difficult to find evidence of citizen responsiveness to performance information. If individuals have access to a continuous set of performance information, signals that come from arbitrary breaks in that set of performance information could be meaningless and, as such, would have little impact on individual behavior (Ahn and Vigdor 2014b). Simply put, discrete school failure signals might add no new information to a fully informed citizenry. In contrast, this shouldn't be a problem if citizens pay limited attention except when an alarm is sounded that government performance has fallen below a particular threshold or tipping point. Under limited information, exogenous school performance signals should enhance citizen learning—giving citizens a clear and accessible marker for what exactly constitutes unacceptable government performance. In short, whether citizens respond to school performance signals is unclear, deserving empirical attention, which the RDD approach used here allows.

METHODS: WHO RESPONDS

The regression discontinuity model used, despite its internal validity, reveals very little about who responds to performance signals. In order to explore this equally important topic, I use two complimentary approaches: model stratification and quantile regression. Both of these have limitations, but they combine to give us a picture of who responds to school failure signals. The first approach simply looks for differences in the regression discontinuity estimates across individual attributes of interest (vote propensity in this case). This approach makes arbitrary decisions about who belongs to high propensity and low propensity groups, estimates an average treatment effect for both groups, and then compares the resultant coefficients. With this approach, when turnout is the outcome, I stratify on whether individuals voted in the previous school board election; when exit is the outcome, I stratify on individuals' SES (from the school records), race/ethnicity (school records), and turnout within the school zone.²⁸ These

²⁵ The exact interval used by the state is not made public and some provisions for NCLB's other academic indicators are somewhat unclear (Ahn and Vigdor 2014a).

²⁶ The standard *F* statistics for instrument strength are provided in the results tables below. The first stage plots are shown in the Online Appendix (see Section VIII).

²⁷ Models are run at the school level and weighted according to the number of registered voters (when turnout is the outcome) or students in the school (when exit is the outcome). This approach preserves the rich nature of the individual-level data, while improving computation time.

²⁸ I use these because I do not have an individual-level match between individual exiters and voters. As such, I have to use

TABLE 1. School Failure and Turnout

	[1] DV: Prop. Voted (Closest)	[2] DV: Prop. Voted (Elementary)	[3] DV: Prop. Voted (Middle)	[4] DV: Prop. Voted (High)
Failure Signal (β_1)	0.025*	0.052*	0.079*	0.051*
	[0.003, 0.047]	[0.021, 0.083]	[0.027, 0.132]	[0.001, 0.102]
Controls	Yes	Yes	Yes	Yes
Fixed Effect (school)	Yes	Yes	Yes	Yes
$F(H_0: \text{weak instr.})$	10,682,588	8,718,299	7,464,324	8,325,778
R^2	0.033	0.045	0.069	0.012
γ	23	145	81	16
N (voter years)	15,220,029	15,928,665	15,372,842	13,991,590
N (school years)	13,561	7,662	3,623	3,141

Notes: * $p < 0.05$. 95% Confidence intervals in braces. Constant suppressed so as to include all fixed effects. Standard errors are clustered to school level. *Unit of analysis*: school-year (population weighted). *Bandwidth*: full range. *Running variable*: quartic. *Controls*: % African American (voter file), % Female (voter file), school size (school file). The Esarey-Danneman parameter which nullifies the estimates is denoted by γ .

variables are informative of what types of voters—be they high or low propensity—react to failure signals.²⁹

To go one step further, when turnout is the outcome I also use quantile regression.³⁰ Quantile regression is preferred by many as a means of exploring treatment heterogeneity because this approach avoids arbitrary decisions of how to define subgroups and because it gives a more comprehensive picture of treatment effects (Angrist and Pischke 2008, Chap. 7).³¹ This alternate approach is a data-driven way to look for treatment heterogeneity. Quantile regression estimates effects across levels of a continuous dependent variable (Koenker 2005; Yu, Lu, and Stander 2003). Put differently, quantile regression examines the effect of treatment on the conditional quantiles of the dependent variable. Given its virtues, quantile regression has been used in a number of situations: as the standard approach to looking for unequal impacts of public works projects (e.g., Gamper-Rabindran, Khan, and Timmins 2010), healthcare interventions (e.g., Austin et al. 2005), education policies (e.g., Eide and Showalter 1998), and educational attainment's effect on income inequality (e.g., Martins and Pereira 2004).³² The use of quantile regression in the NCLB application is similar. Here, I combine quantile regression with the regression discontinuity models to see whether school failure signals increase turnout around high turnout schools more than low turnout schools, or, put differently, whether failure signals shift the top of the school turnout distribution more than the bottom. This

approach preserves the internal validity of the regression discontinuity models, while giving us a data-driven approach to examine what type of citizens respond to school failure signals. If failure signals have larger effects as we move up the turnout distribution, we can conclude these promote a citizen response most among high propensity citizens.³³

RESULTS

I first examine how failure signals influence voice-based citizen responses. Table 1 shows the effect of school performance signals on voter turnout. The results indicate that having a marginally failing school nearby increases school board turnout by approximately five to eight percentage points. The estimates are highly significant regardless of the school level considered.³⁴ Signaled failure, in short, increases voter turnout in school board elections.

When considering the substantive meaning of these effects, a few comparisons are useful. First, we can compare these estimates to the turnout distribution. When standardizing the outcome we can see that failure's effect in closest (0.09σ), elementary (0.19σ), middle (0.31σ), and high (0.20σ) schools is noticeable.³⁵ Second, we can compare the estimates to the margin of victory in school board elections. Over the period of study in North Carolina, school board elections were generally very close. When compared to the margin of victory, the turnout increases are greater than or equal to

information that is available at the individual level in the school files or at the school level in the voter files that proxy for turnout.

²⁹ In North Carolina, the race/ethnicity and SES participation gaps are noticeable, with both factors being strongly predictive of turnout.

³⁰ Quantile regression is less appropriate when exit is the outcome because the exit distribution may not correspond with vote propensity.

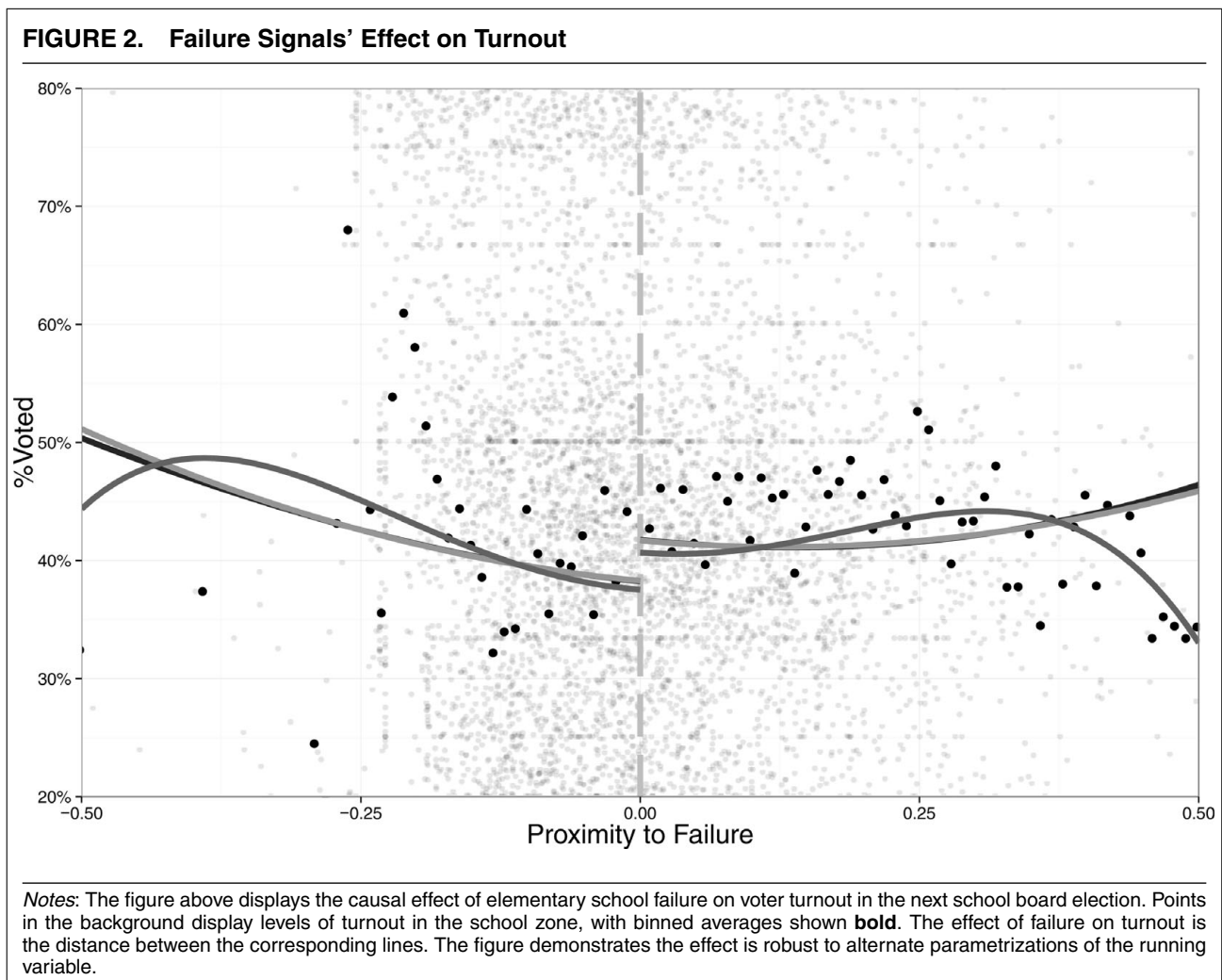
³¹ To use quantile regression the distribution of the outcome must be continuous and relatively smooth across the distribution. Turnout in school zones satisfies these conditions.

³² In this application, the question remains whether the causal impact of attainment has been identified.

³³ School fixed effects are omitted in the quantile models, as combining these two is methodologically problematic (Gamper-Rabindran, Khan, and Timmins 2010).

³⁴ This is true even with a conservative approach to estimating the standard errors. Here I cluster the standard errors at a higher level than treatment (the school instead of the school-year level). This is done due to the inherent difficulties involving non-nested data with fixed effects.

³⁵ The limitations of standardized coefficients are well known. I include these numbers as a reference for education policy researchers, who frequently use these.



31% (closest), 43% (elementary), 48% (middle), and 43% (high) of the school board races. Third, using the strong body of work from GOTV studies as a benchmark, school failure compares favorably. The mobilizing effect of a failing school signal is smaller than a typical face-to-face contact, but noticeably larger than typical phone interventions and mailers (Green and Gerber 2008; Green, Gerber, and Nickerson 2003). Finally, we can implement a Bayesian technique that estimates a parameter for how risk averse to false positives a researcher would have to be in order to nullify a statistically significant effect (Esarey and Danneman 2014).³⁶ When applied to the school failure estimates, this technique shows that even with a very strong aversion to false positives, the effect of a failing school on turnout remains robust. Thus, regardless of the technique used, school failure's mobilizing effect appears substantively large.

Figure 2 shows this effect visually (see also Figure A.5 in the Online Appendix). Failure's effect on

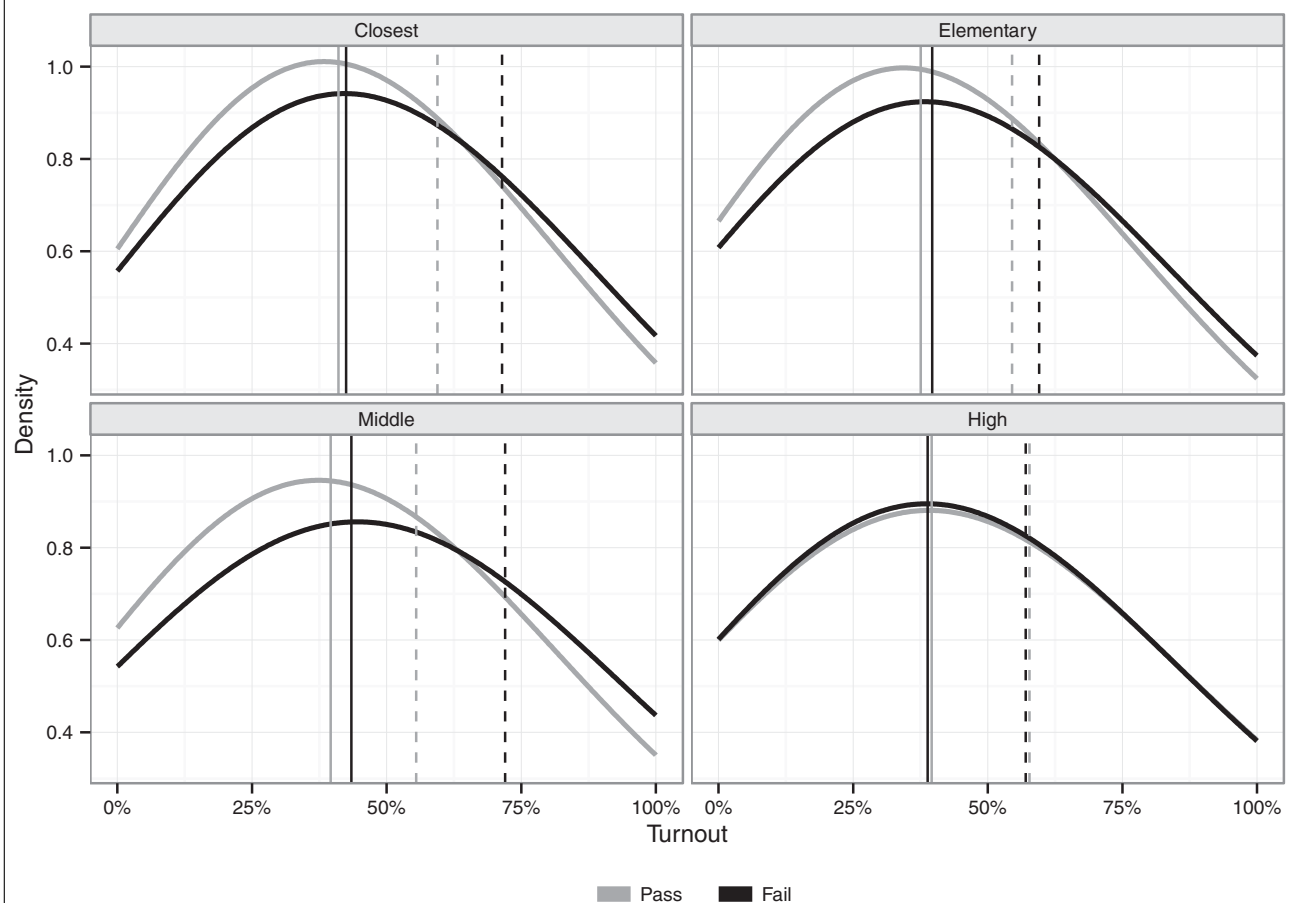
turnout can be seen in the vertical jump in turnout at the failure cutoff. This effect is apparent across various alternate specifications of the running variable.

These results show that performance information increases turnout. Still, the possibility remains that a particularly active group drives this increase. As mentioned earlier, existing theories offer conflicting predictions for how responsiveness to performance information varies across individuals' propensity to participate. If school failure mobilizes high- and low-turnout individuals equivalently, we should expect to see that the results are similar across models that stratify on vote history and across the coefficients of the quantile regressions.

In reality, this is not what we observe. Instead, failure signals are much more likely to mobilize people who have previously voted in school board elections. The treatment effect for those who voted in the last election is 4.6 percentage points ($p \approx 0.02$), while the treatment effect for those who did not is only 0.8 percentage points ($p \approx 0.67$).³⁷ These coefficients are substantively

³⁶ This process estimates a parameter that describes the valuation ratio for correct and incorrect decisions. Esarey and Danneman recommend a ratio of at least $\gamma = 2$ for substantive significance (2014, 14).

³⁷ The same pattern holds when stratifying on individual turnout two (voted: 2.2%; not voted: 0.0%) and three elections previous (voted: 5.2%; not voted: 1.3%).

FIGURE 3. Failure Signals and Participatory Inequality

Notes: The two distributions are smoothed kernel densities with a bandwidth of 0.1 points around the failure cutoff. The lines represent the 50th (solid) and 75th percentiles (dashed). The distance between the corresponding black and grey lines represent coefficient estimates from 50th and 75th quantile regressions. The figures show that failure signals move the top of the turnout distribution more than the bottom, showing an unequal response.

different (the effect for previous voters is 5.75 times larger than the effect for nonvoters) and statistically distinct ($p < 0.05$). The quantile regressions show the same thing. From these we can see that failure moves the top of the turnout distribution more than the bottom. For example, school failure raises the 80th percentile of the turnout distribution by 13% while only moving turnout at the 10th percentile by 5%. Like the stratification results, these two coefficients are statistically and substantively distinct.

Figure 3 shows the quantile regression results visually, plotting the turnout distribution for marginally passing and marginally failing schools. Overlaid are lines for the 50th (solid) and 75th (dashed) percentiles, corresponding to both distributions. The difference between these corresponding percentile lines represents the quantile regression coefficients at the 50th and 75th percentiles. As can be seen, failure primarily shifts the upper portion of the turnout distribution. Comparing groups in the bottom turnout quartile and those in the top reveals a treatment effect that is 5.5 percentage

points greater for high turnout individuals.³⁸ (Figure A.9 in the Online Appendix provides a more detailed plot of the quantile regression coefficients, with their corresponding levels of uncertainty.)

What do these results mean? Because school failure signals predominantly mobilize those who have recently voted and those at the top of turnout distribution, we can infer that performance information primarily promotes participation among those that are already likely to engage at baseline. In short, this evidence is consistent with performance signals encouraging an unequal response. This suggests that an important stratification not previously considered in accountability models. Performance information may

³⁸ The differences between the 25th and 75th quantile estimates in closest and elementary schools are both substantively noticeable and statistically meaningful (closest, $p = 0.03$; elementary, $p = 0.03$). In middle schools, the difference is substantively meaningful, but not precise enough—perhaps due to the relatively few number of middle schools—to drawn clear inferences ($p = 0.47$). In high schools, the pattern is less clear ($p = 0.22$).

increase average levels of citizen engagement, but it may do so at the cost of promoting participatory inequality.

Beyond changing turnout levels, failure signals appear to also influence election outcomes. To see this, we must step back from the level of the citizen—due to the secret nature of the ballot—to a higher level of aggregation. Such an approach models the margin of victory and the number of candidates running in school board elections as a function of failure and proximity to failure in school districts. I specify treatment in these models through a count measure of the number of schools failing in the district in a given year. For these models, proximity is calculated using the average running variable score in the district. Even at this higher level of aggregation, there exists a discrete jump in the number of failing schools at the point where the average district score crosses zero. As such, the results are analogous to the school-level RDD models.³⁹ The results from these models can be interpreted as estimates of how citizens respond when school failure in a given school district increases.

Table 2 shows the results of these models. The results suggest that when citizens receive signals that schools in their district are performing poorly, they hold local school board officials accountable. Races with more failing schools see more candidates running and tighter election outcomes. When the number of elementary schools failing increases by a standard deviation (≈ 6 schools), school board races see about one to two additional candidates running and margins of victory that are about one to three points narrower. This shift is noticeable: large enough to swing about 25% of school board elections observed in the sample. Put differently, a simple calculation suggests that somewhere between 20 and 60% of the citizens mobilized by failure altered their vote choice.⁴⁰ In short, when schools fail, citizens become willing for a change in the school board leadership. Potential challengers, noticing the opportunity, choose to run at a higher rate.

Qualitative evidence is consistent these empirical findings. Wake County (home of Raleigh—the second largest city in the state) Public Schools' experience is illustrative. In 2008, 80% of schools failed. This put residents in this model district in unfamiliar territory—in the bottom quartile of the state. This large increase in the number of failing schools shaped the number of challengers that ran for the board—an abnormally high nine ran in the next election to fill four seats, putting the race in the top quartile of race competitiveness. Christopher Malone, a school board challenger—and eventual board member—stated bluntly that he was challenging for a spot on the school board because, “we have too many failing schools [in Wake County].” Likewise, after a large number of Chatham County schools received failing marks in 2007, the subsequent school

board race was particularly contentious—eliciting a retirement and two challengers for a seat that had previously gone unchallenged. The race came down to a 2.5 percentage point margin—three times smaller than the surrounding elections in the district. Before the election, challenger Flint O'Brien described his reasons for running as fulfilling a desire to “work to fix [Chatham's failing schools] instead of running away,” alluding to the option many parents were taking to exit failing schools in the district.^{41,42}

RESULTS: FAILURE SIGNALS AND EXIT

Besides working to remove those deemed responsible for failure, citizens may choose to exit failing schools: voting with their feet instead of at the ballot box. To examine whether this occurs, I turn to data on the number of students exiting individual schools each year.

My regression discontinuity models indicate that school failure signals also cause an increase in residents' voting with their feet. Table 3 shows that school failure causes residents to exit, causing about 16 more families to leave in the next year than in marginally passing schools. Figure 4 presents this effect visually, by showing the discrete jump in exit at the failure cut-off. Of the 16 failure-induced exits, about 10 families transfer to other schools, most of which occur within districts.⁴³ These estimates are all highly significant. Citizens, in short, use exit when they are given a signal of low public school performance.

To consider the substantive significance of these estimates, a few comparisons are illuminating. First, we can compare these estimates to the distribution of exit from schools. The results represent about 10% of a standard deviation of the usual exits that occur in North Carolina schools in a given year—a noticeable amount. Second, when we compare the coefficients to the distribution of school size in North Carolina schools, it can be seen that the exit estimate represents just less than 1% of an average school's population. Though this may seem small, this effect is meaningful. The number of students exiting in response to failure means the loss revenue for

⁴¹ See Chris Malone and Flint O'Brien candidate questionnaires at www.indyweek.com.

⁴² Some incumbents—perhaps anticipating challengers—also seem conscientious of NCLB labeling their schools failing. In 2011, when only four of 27 schools in the Nash-Rocky Mount Public Schools made AYP, Evelyn Bulluck—board vice-chairwoman, who was up for re-election the next year—remarked “Proficiency levels are way down, and we're not pleased with those schools that haven't made AYP . . . We are displeased at having not met our standards. We will need to come up with a different strategy to improve” (see “Nash Schools Need Improvement this Year,” *The Bailey/Middlesex News*, 2011).

⁴³ Transfers are allowed if a school has failed twice consecutively and is under the first year of sanctions. Most of the failure-induced transfers occur when this is the case. However, a few transfers out of schools do occur when a school fails and is not under sanction. This estimate is about a quarter of the size of the estimate when sanctions are in place, and is statistically significant. These nonsanction transfers are possible because the exemptions offered—general, hardship, or safety exemptions—are available and because of the state's growing number of charter schools.

³⁹ These models omit the school fixed effect because the level of observation is now the district and precise manipulation is much less likely at this higher level of aggregation.

⁴⁰ ($1\% \Delta \text{MOV} / 5\% \Delta \text{turnout} = 20\%$) or ($3\% \Delta \text{MOV} / 5\% \Delta \text{turnout} = 60\%$).

TABLE 2. Failure Signals and Electoral Competition

	[1] DV: #Running/Seats (Elementary)	[2] DV: #Running/Seats (Elementary)	[3] DV: #Running/Seats (Middle)	[4] DV: #Running/Seats (Middle)	[5] DV: #Running/Seats (High)	[6] DV: #Running/Seats (High)
Failure Signal	1.37* [1.27, 1.46]	1.42* [1.23, 1.62]	1.33* [1.23, 1.42]	1.76* [1.53, 1.98]	1.05* [.933, 1.17]	-0.492* [-0.724, -0.259]
Controls	No	Yes	No	Yes	No	Yes
Constant	2.81* [2.63, 2.99]	3.90* [3.50, 4.30]	2.76* [2.58, 2.95]	4.19* [3.78, 4.60]	2.70* [2.49, 2.91]	1.91* [1.48, 2.33]
R ²	0.445	0.492	0.424	0.497	0.224	0.395
γ	> 100	> 100	> 100	> 100	> 100	> 100
N	1,062	1,062	1,062	1,062	1,062	1,062
	[7] DV: Margin of Victory (Elementary)	[8] DV: Margin of Victory (Elementary)	[9] DV: Margin of Victory (Middle)	[10] DV: Margin of Victory (Middle)	[11] DV: Margin of Victory (High)	[12] DV: Margin of Victory (High)
Failure Signal	-0.017* [-0.028, -0.007]	-0.025* [-0.044, -0.006]	-0.015* [-0.024, -0.005]	-0.037* [-0.061, -0.014]	-0.007 [-0.017, 0.003]	0.003 [-0.018, 0.026]
Controls	No	Yes	No	Yes	No	Yes
Constant	0.118* [0.095, 0.142]	0.400* [0.340, 0.458]	0.119* [0.096, 0.143]	0.387* [0.328, 0.446]	0.119* [0.096, 0.144]	0.429* [0.368, 0.490]
R ²	0.020	0.285	0.018	0.224	0.013	0.283
γ	> 100	41	> 100	100	7	2
N	1,064	1,064	1,064	1,064	1,064	1,064

Notes: * $p < 0.05$. 95% Confidence intervals in braces; robust standard errors. MOV = margin of victory. *Unit of analysis*: school district (candidate weighted). *Bandwidth*: full range. *Running variable*: quartic. *Controls*: # of citizens, # vote for, # of schools, % of students Af. American. γ is the Esarey-Danneman risk-aversion parameter.

TABLE 3. Failure Signals and Exit

	[1] DV: # Exit	[2] DV: # White Exit	[3] DV: # Black Exit	[4] DV: # Hispanic Exit	[5] DV: # Affluent Exit	[6] DV: # Non Affluent Exit	[7] DV: # Exit (High Turnout)	[8] DV: # Exit (Low Turnout)
Fail Signal (β_1)	15.99* [1.22, 30.78] Yes	10.46* [3.61, 17.30] Yes	2.44 [-6.08, 10.95] Yes	2.16 [-0.330, 4.64] Yes	28.4* [5.09, 51.72] Yes	6.39 [-6.84, 19.62] Yes	16.85* [0.604, 33.09] Yes	7.25 [-3.90, 18.39] Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	2,288,074	2,288,074	2,288,074	2,288,074	2,288,074	2,288,074	2,288,074	2,288,074
F-Weak Instr.	0.15	0.117	0.113	0.119	0.206	0.230	0.969	0.942
R ²	19	88	3	10	30	4	17	6
γ	4,914,921	4,914,921	4,914,921	4,914,921	4,914,921	4,914,921	1,425,972	3,547,646
N (students)	10,236	10,236	10,236	10,236	10,236	10,236	2,894	7,342
N (school years)								

Notes. * $p < 0.05$. 95% Confidence intervals in braces. Constant suppressed so as to include all fixed effects. *Unit of analysis:* school year (weighted). Standard errors are clustered to school level. *Bandwidth:* full range. *Running variable:* quartic. *Controls:* % African American (voter file), % Female (voter file), school size (school file). γ is the Esarey-Danneman risk-aversion parameter (recommended threshold: $\gamma = 2$). Affluence measured through free-reduced price lunch.

schools. Given the state’s funding formula, the number of students lost is approximately equal to one classroom (or one teacher). For public schools this also means the loss of funding for other programs. Taken together, a back-of-the-envelope calculation reveals that failure-induced exit leads to about a \$54,000 decline in funding.⁴⁴ This amount, though perhaps too small for the general population to notice, represents a crucial amount for school officials, especially in times of tight education budgets.⁴⁵

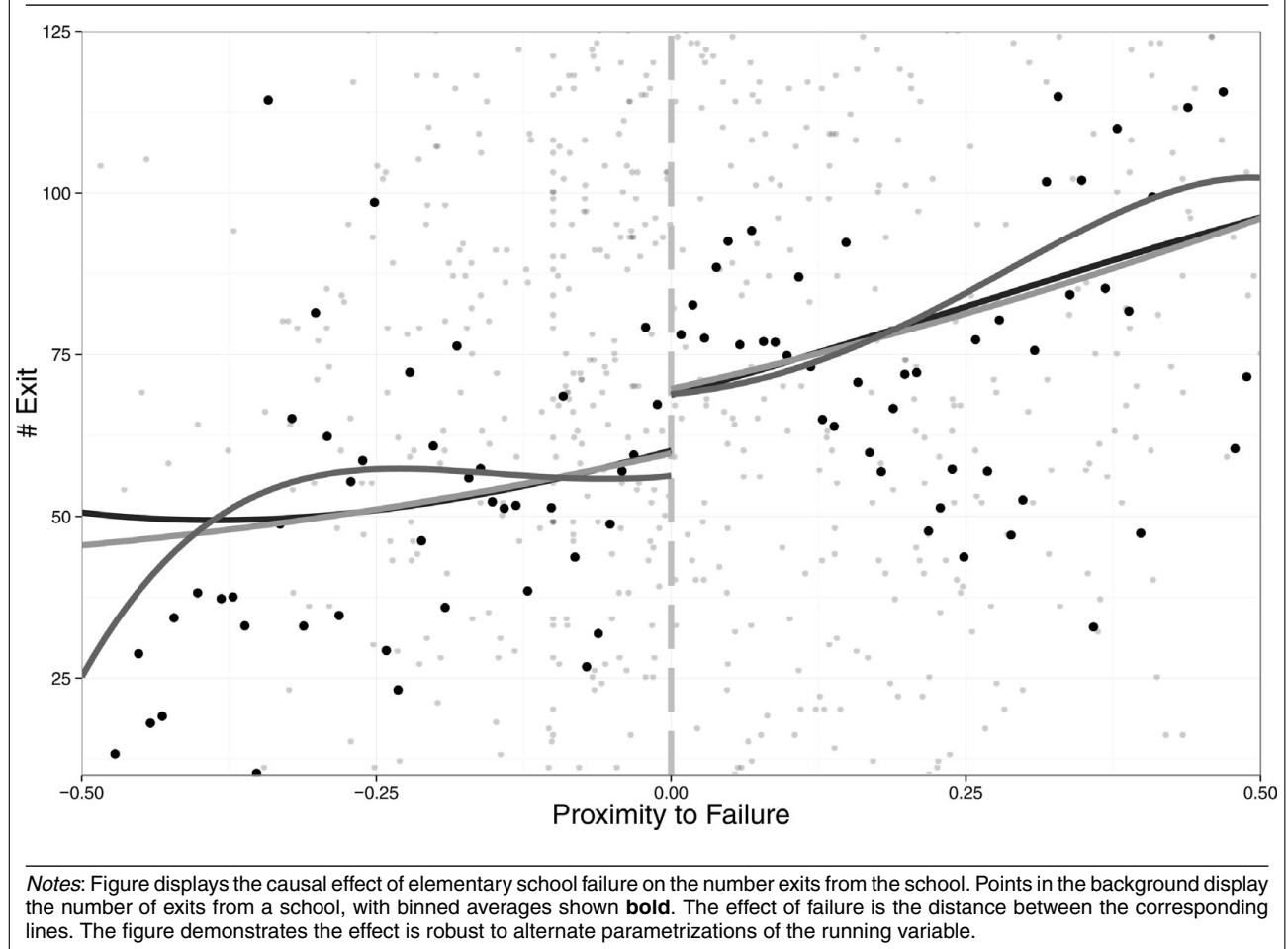
In short, it appears that failure signals—in addition to shifting patterns of citizen voice—also have a meaningful effect on patterns of citizen exit. This result is important; it suggests that not only does performance information spur voice-based responses—mobilizing and shifting voters’ choices, as others have shown—it also promotes citizen responsiveness through exit.

That said, like voice-based responses, only a select few use exit. Table 3 shows this by presenting the coefficients stratified over three proxies of vote propensity: socioeconomic status, race/ethnicity, and the turnout in the area around the school. If failure signals encourage exit among low propensity residents, we would expect to see low SES individuals, minorities, and those living in low turnout areas to use exit when their school fails.

Yet, this is not what we observe. Columns 1–3 first explore racial patterns in failure-induced exit. These show that school failure causes white residents to exit much more than minorities. Columns 4 and 5 show exit patterns across levels of affluence. These indicate that affluent students are more likely than poor to exit in response to school failure signals. Finally, columns 6 and 7 show exit patterns across school turnout levels. These models indicate that high turnout areas are more likely to see higher exit in response to a failure signal than low. In short, failure is most likely to elicit an exit response from white, affluent families in high turnout school zones. Together, these models suggest that performance information does not empower low propensity participators, as some have argued (e.g., Warren 2011), through exit. This distinction is meaningful, suggesting that rather than being an empowering alternate venue for citizen accountability, exit does little to fill the voice-based gaps caused by performance signals.

⁴⁴ This estimate was based on the cost of a teacher and the cost of other programs. Failure-induced exit amounts to a loss of about one teacher. In the state, teacher salaries range about \$30–\$45k per year. If we use the lower salary estimate, the funding lost for teachers due to failure is \$30k. Regarding other programs, schools are allocated about \$1,500 per student for various programs (teacher assistants, etc.). Thus, schools that lose 16 students due to failure receive about \$24,000 less for these programs. Hence: \$30,000 (teacher salary) + \$24,000 (other programs).

⁴⁵ Such post-treatment declines in funding are unlikely to drive the voice responses documented here. I explore this possibility empirically in the Online Appendix. In short, changes in funding that do occur, but likely do after the board elections are held. More importantly, failure-induced changes to funding are not signaled to the public in a transparent way and are relatively small compared to the broader funding structure that is locked into place before school failure status is determined.

FIGURE 4. Failure Signals' Effect on Exit

CONCLUSION

This article provides causal evidence that citizens are responsive to government performance information provided through performance accountability policies. I have shown that when school performance falls below No Child Left Behind's failure threshold, citizens voice their displeasure, turning out at higher rates in school board elections and increasing their competitiveness. Additionally, as theory has long predicted but few studies have explored, performance information also primes citizens' use of exit. Given these increases in citizen responsiveness, performance accountability appears to have a role to play in remediating low levels of citizen information. However, these gains do not distribute themselves equally. With both voice and exit, high propensity citizens appear to be much more likely to react to a signal than low propensity citizens. Instead of performance accountability reforms filling gaps in citizen responsiveness, as many of them intend, these appear to actually exacerbate inequalities in citizen responsiveness.

Given citizen responsiveness' key role in models of democratic accountability, these offer two key contributions to research in this area. First, the results suggest

that accountability studies have focused myopically on votes and other voice-related metrics. It appears that citizens use exit as an alternate venue when they are dissatisfied with government performance. As such, those who fail to explore exit in accountability studies may come to the wrong conclusions about citizen responsiveness to information about government performance. Given exit's potential as a tool for enhancing accountability, this act has been underexplored.⁴⁶ Second, the results suggest that contemporary accountability models should consider the possibility that citizens respond unequally to the same information about government performance. Models that look only for a uniform response may miss meaningful heterogeneity in citizen responsiveness. This finding has important policy implications. If we believe that citizen responsiveness drives government accountability, the likely result of performance signals is to encourage inequalities in democratic accountability.⁴⁷ Given these results,

⁴⁶ It remains to be seen whether the theoretical prediction that elected officials actually respond to exit is actually realized. Such a topic is valuable, but simply beyond the scope of this article.

⁴⁷ Of course, responsiveness from elected officials may break down (e.g., Ashworth 2012; Healy and Malhotra 2009; 2013; Nielsen and

performance accountability approaches to addressing information gaps may not enhance accountability for all.

The source of this unequal response is unclear. If this inequality arises because less sophisticated, low propensity citizens do not have easy access to performance information, policy enhancements regarding how performance information is formatted and distributed may be in order. If, however, these gaps arise because low propensity citizens have more obstacles in the way of exercising voice and exit, changes to performance information may not be enough to close this gap. For example, if low propensity participants would like to use exit to leave failing schools, but have limited capacity to do so—because of moving or transportation costs—performance accountability reforms may not be enough. Such a situation would call for complimentary policies that simultaneously target the elimination of these obstacles.

Regardless of the reasons driving this unequal response, this finding has troubling implications for performance accountability reforms. Performance accountability systems are widespread and rapidly spreading—these reforms can now be found in health, penal, nonprofit, environmental, agricultural, and foreign policy sectors. Many of these systems, including No Child Left Behind, intend to not only draw out a citizen response, but to do so broadly: with these reforms placing special focus on empowering disadvantaged citizens. The results presented here suggest that these systems may struggle to achieve this objective, and may even make things worse. Thus, policy-makers should proceed with care—these systems may spur local accountability among some, but leave many citizens perpetually behind.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S0003055416000071>.

REFERENCES

- Ahn, Thomas and Jacob Vigdor. 2014a. “The Impact of No Child Left Behind’s Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina.”; National Bureau of Economic Research Working Paper No. w20511.
- Ahn, Thomas, and Jacob L. Vigdor. 2014b. “When Incentives Matter Too Much: Explaining Significant Responses to Irrelevant Information.” National Bureau of Economic Research Working Paper No. w20321.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press.
- Arceneaux, Kevin, and David W. Nickerson. 2009. “Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments.” *American Journal of Political Science* 53 (1): 1–16.
- Arnold, R. Douglas, and Nicholas Carnes. 2012. “Holding Mayors Accountable: New York’s Executives from Koch to Bloomberg.” *American Journal of Political Science* 56 (4): 949–63.

Moynihan 2015). Future research should pay careful attention to the effect of responsiveness inequalities on elected official responses, governance quality, and citizen welfare. In particular, substantial work remains to see the downstream consequences of unequal exit.

- Ashworth, Scott. 2012. “Electoral Accountability: Recent Theoretical and Empirical Work.” *Annual Review of Political Science* 15: 183–201.
- Austin, Peter C., Jack V. Tu, Paul A. Daly, and David A. Alter. 2005. “The Use of Quantile Regression in Health Care Research: A Case Study Examining Gender Differences in the Timeliness of Thrombolytic Therapy.” *Statistics in Medicine* 24 (5): 791–816.
- Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. “Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India.” *American Economic Journal: Economic Policy* 2 (1): 1–30.
- Barrows, Samuel. 2014. “Performance Information and Retrospective Voting: Evidence from a School Accountability Regime.” Working Paper.
- Berry, Christopher R., and William G. Howell. 2007. “Accountability and Local Elections: Rethinking Retrospective Voting.” *Journal of Politics* 69 (3): 844–58.
- Björkman, Martina, and Jakob Svensson. 2009. “Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda.” *The Quarterly Journal of Economics* 124 (2): 735–769.
- Butler, Daniel M., and Matthew J. Butler. 2006. “Splitting the Difference? Causal Inference and Theories of Split-Party Delegations.” *Political Analysis* 14 (4): 439–55.
- Campbell, Angus, Phillip E. Converse, Warren E. Miller, and Donald E. Stokes. 1980. *The American Voter*. Chicago: University of Chicago Press.
- Chen, Jowei. 2013. “Voter Partisanship and the Effect of Distributive Spending on Political Participation.” *American Journal of Political Science* 57 (1): 200–17.
- Chingos, Matthew M., Michael Henderson, and Martin R. West. 2012. “Citizen Perceptions of Government Service Quality: Evidence from Public Schools.” *Quarterly Journal of Political Science* 7 (4): 411–45.
- Chong, Alberto, and Ana L. De La O., Dean Karlan, and Leonard Wantchekon. 2011. “Looking Beyond the Incumbent: The Effects of Exposing Corruption on Electoral Outcomes.” National Bureau of Economic Research Working Paper No. w17679.
- Chubb, John E., and Terry M. Moe. 1991. *Politics, Markets, and America’s Schools*. Washington, DC: Brookings Institution Press.
- Cox, James H., and Christopher Witko. 2008. “School Choice and the Creation of Social Capital Reexamined.” *American Journal of Political Science* 52 (1): 142–55.
- Dee, Thomas S., and Brian Jacob. 2011. “The Impact of No Child Left Behind on Student Achievement.” *Journal of Policy Analysis and Management* 30 (3): 418–46.
- Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know About Politics and Why it Matters*. New Haven, CT: Yale University Press.
- Di Gennaro, Corinna, and William Dutton. 2006. “The Internet and the Public: Online and Offline Political Participation in the United Kingdom.” *Parliamentary Affairs* 59 (2): 299–313.
- Dowding, Keith, and Peter John. 2008. “The Three Exit, Three Voice and Loyalty Framework: A Test with Survey Data on Local Services.” *Political Studies* 56 (2): 288–311.
- Eide, Eric, and Mark H. Showalter. 1998. “The Effect of School Quality on Student Performance: A Quantile Regression Approach.” *Economics Letters* 58 (3): 345–50.
- Enos, Ryan D., Anthony Fowler, and Lynn Vavreck. 2014. “Increasing Inequality: The Effect of GOTV mobilization on the composition of the electorate.” *The Journal of Politics* 76 (01): 273–88.
- Esarey, Justin, and Nathan Danneman. 2014. A Quantitative Method for Substantive Robustness Assessment. Political Science Research and Methods, available on CJO2014. doi:10.1017/psrm.2014.14.
- ESEA. 2002. “No Child Left Behind Act of 2001.”
- Ferejohn, John. 1986. “Incumbent Performance and Electoral Control.” *Public Choice* 50 (1): 5–25.
- 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): 703–45.
- Figlio, David N., and Maurice E. Lucas. 2004. “What’s in a Grade? School Report Cards and the Housing Market.” *American Economic Review* 94 (3): 591–604.

- Fiorina, Morris P. 1978. "Economic Retrospective Voting in American National Elections: A Micro-Analysis." *American Journal of Political Science* 22 (2): 426–43.
- Friedman, Milton. 1955. *The Role of Government in Education*. New Brunswick, NJ: Rutgers University Press.
- Gamper-Rabindran, Shanti, Shakeeb Khan, and Christopher Timmins. 2010. "The Impact of Piped Water Provision on Infant Mortality in Brazil: A Quantile Panel Data Approach." *Journal of Development Economics* 92 (2): 188–200.
- Grant, Ruth W., and Robert O. Keohane. 2005. "Accountability and Abuses of Power in World Politics." *American Political Science Review* 99 (01): 29–43.
- Green, Donald P., and Alan S. Gerber. 2008. *Get out the Vote: How to Increase Voter Turnout*. Washington, DC: Brookings Institution Press.
- Green, Donald P., Alan S. Gerber, and David W. Nickerson. 2003. "Getting out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments." *Journal of Politics* 65 (4): 1083–96.
- Grose, Christian R., and Bruce I. Oppenheimer. 2007. "The Iraq War, Partisanship, and Candidate Attributes: Variation in Partisan Swing in the 2006 US House Elections." *Legislative Studies Quarterly* 32 (4): 531–57.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design." *Econometrica* 69 (1): 201–9.
- Healy, Andrew, and Neil Malhotra. 2009. "Myopic Voters and Natural Disaster Policy." *American Political Science Review* 103 (03): 387–406.
- Healy, Andrew, and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16: 285–306.
- Helliwell, John F., and Robert D. Putnam. 2007. "Education and Social Capital." *Eastern Economic Journal* 33 (1): 1–19.
- Henderson, Michael. 2010. "Does Information Help Families Choose Schools? Evidence from a Regression Discontinuity Design." Working paper.
- Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Vol. 25. Cambridge, MA: Harvard University Press.
- Holbein, John, and Helen F. Ladd. 2015. "Accountability Pressure and Non-Achievement Student Behaviors." Calder Working Paper No. 122.
- Holbein, John B., and D. Sunshine Hillygus. forthcoming. "Making Young Voters: The Impact of Preregistration on Youth Turnout." *American Journal of Political Science*.
- Holbrook, Thomas M. 2002. "Presidential Campaigns and the Knowledge Gap." *Political Communication* 19 (4): 437–54.
- Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.
- Jacob, Brian A., and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics* 86 (1): 226–44.
- Jeynes, William H. 2007. "The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement A Meta-Analysis." *Urban Education* 42 (1): 82–110.
- Koenker, Roger. 2005. *Quantile Regression*. Cambridge, United Kingdom: Cambridge University Press.
- Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. forthcoming. "Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda." *American Journal of Political Science*.
- Lassen, David Dreyer. 2005. "The Effect of Information on Voter Turnout: Evidence from a Natural Experiment." *American Journal of Political Science* 49 (1): 103–18.
- Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48 (2): 281–355.
- Lewis-Faupel, Sean, Yusuf Neggars, Benjamin A. Olken, and Rohini Pande. 2014. "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia." National Bureau of Economic Research Working Paper No. w20344.
- Martins, Pedro S., and Pedro T. Pereira. 2004. "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries." *Labour Economics* 11 (3): 355–71.
- Matsudaira, Jordan D. 2008. "Mandatory Summer School and Student Achievement." *Journal of Econometrics* 142 (2): 829–50.
- McDonnell, Lorraine M. 2009. "Repositioning Politics in Education's Circle of Knowledge." *Educational Researcher* 38 (6): 417–27.
- Minkoff, Scott. 2014. "Political Engagement in a Public Goods Context: The Effect of Neighborhood Conditions and Schools on Local Election Turnout." SSRN Working Paper No. 2416914
- Nielsen, Poul A., and Donald P. Moynihan. 2015. How Do Elected Officials Attribute Responsibility for Government Performance? Negativity, Advocacy, and Partisan Responses. Working paper.
- Niemi, Richard G., Harold W. Stanley, and Ronald J. Vogel. 1995. "State Economies and State Taxes: Do Voters Hold Governors Accountable?" *American Journal of Political Science* 29 (4): 936–57.
- Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.
- Pande, Rohini. 2011. "Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies." *Annual Review of Economics* 3 (1): 215–37.
- Papay, John P., John B. Willett, and Richard J. Murnane. 2011. "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables." *Journal of Econometrics* 161 (2): 203–7.
- Przeworski, Adam, Susan C. Stokes, and Bernard Manin, eds. 1999. *Democracy, Accountability, and Representation*. Vol. 2. Cambridge, United Kingdom: Cambridge University Press.
- Rhodes, Jesse H. 2014. "Learning Citizenship? How State Education Reforms Affect Parents' Political Attitudes And Behavior." *Political Behavior* 37 (1): 181–220.
- Rich, Peter M., and Jennifer L. Jennings. forthcoming. "Choice, Information, and Constrained Options School Transfers in a Stratified Educational System." *American Sociological Review*.
- Rudolph, Thomas J. 2003. "Who's Responsible for the Economy? The Formation and Consequences of Responsibility Attributions." *American Journal of Political Science* 47 (4): 698–713.
- Schlozman, Sidney Verba, Kay Lehman, and Henry E. Brady. 2012. *The Uneheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy*. Princeton, NJ: Princeton University Press.
- Stecher, Brian M., Frank Camm, Cheryl L. Damberg, Laura S. Hamilton, Kathleen J. Mullen, Christopher Nelson, and Paul Sorensen. 2010. *Toward a Culture of Consequences*. Santa Monica, CA: Rand Corporation.
- Stock, James H., and Motohiro Yogo. 2005. Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Eds. Donald W. K. Andrews, James H. Stock, and Thomas J. Rothenberg.
- Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *The Journal of Political Economy* 43 (5): 416–24.
- Trounstein, Jessica. 2010. "Representation and Accountability in Cities." *Annual Review of Political Science* 13: 407–23.
- Warren, Mark E. 2011. "Voting with Your Feet: Exit-Based Empowerment in Democratic Theory." *American Political Science Review* 105 (04): 683–701.
- Wolfinger, Raymond E., Benjamin Highton, and Megan Mullin. 2005. "How Postregistration Laws Affect the Turnout of Citizens Registered to Vote." *State Politics & Policy Quarterly* 5 (1): 1–23.
- Yu, Keming, Zudi Lu, and Julian Stander. 2003. "Quantile Regression: Applications and Current Research Areas." *Journal of the Royal Statistical Society: Series D (The Statistician)* 52 (3): 331–50.