When Your Group Fails: The Effect of Race-Based Performance Signals on Citizen Voice and Exit

John B. Holbein
Assistant Professor of Political Science
Brigham Young University
John_holbein@byu.edu

Hans Hassell
Assistant Professor of Political Science
Florida State University
hans.hassell@fsu.edu

Abstract:
In recent years, performance-based accountability reforms have become widespread—particularly in the public education system—as a means of enhancing service delivery, knowledge, and citizen engagement with government. While an ever-growing literature has considered the overall effects of these administrative systems, few have fully considered the effect of such systems on underlying subgroups of relevance and inequalities in responses across these groups. In this paper, we examine how citizens of various racial subgroups respond when governments send negative race-based signals about the performance of in- and out-racial groups. Specifically, we explore whether No Child Left Behind's (NCLB) race-based failure signals affect racial groups' use of voice in school board elections *and* their use of exit from local public schools. To do so, we combine school and voter administrative records in North Carolina with a regression discontinuity design that leverages exogenous variation at NCLB's race-based school failure cutoffs. Consistent with our theoretical framework, we find that white and black citizens respond differently to race-based failure signals. We find the whites are more responsive overall; responding to failure signals of various types both by voting in local school board elections and exiting local failing schools. African Americans, however, seem much more sensitive to racialized cues—responding at the ballot box but *not* by exiting when their group fails. These results show that while performance accountability systems have broader effects than previously realized: shifting the composition of local elections and distorting the racial makeup of schools.

Abstract Word Count: 242
Word Count: 11,327

When signed into law in 2002, No Child Left Behind (NCLB) was touted as a policy that provided choices to students in struggling schools. Decrying the "soft bigotry of low expectations," and as signaled in the law's name, President George W. Bush's signature education policy was particularly aimed at enabling and strengthening education outcomes among low performing minority students. Speaking to the NAACP as a candidate in 2000, Bush explained, "There's a tremendous gap of achievement between rich and poor, white and minority. This, too, leaves a divided society" and argued that part of the solution to the problem was to "provide parents with options to increase their option, like charters and choice" (Bush 2000). Part of this initiative involved evaluating schools' performance on yearly student learning outcomes and publicizing performance information to surrounding communities. This provided citizens with information about the overall performance of schools as well as information about various racial subgroups within those schools. The thought was that signaling low performance would help mobilize parents to use voice to express their dissatisfaction or to leave when school performance deteriorated. In short, NCLB was passed as a means of enabling parents to seek a better education by helping them hold school leaders accountable and allowing flexibility in school choice.

NCLB was an integral part of a larger movement towards promoting accountability in recent years (Moynihan 2008; Moynihan and Ingraham 2003; Stecher et al. 2009; Jacobsen and Saultz 2016; Michener and Ritter 2017; Payson 2017). While governments have long been a source of information for citizens, in recent years, numerous institutions have formalized this role by implementing performance accountability systems (Heinrich 2002; Heinrich 2009; Kroll, Neshkova, and Pandey 2017; Kroll 2017; Radin 2000; Miller 2017; Moynihan, Pandey, and Wright 2011; Nielsen and Moynihan 2017). These systems have also been referred to as

"standards-based accountability," "social accountability," "results-based accountability," and "performance regimes" (Martins et al. 2016). These systems vary, but generally consist of measurement, standards, publication of performance, and/or sanctions for low performance. These have become increasingly common in international development in the form of politician monitoring and audits (e.g., Banerjee et al. 2010; Grossman and Hanlon 2013), in the financial sector in the form of disclosure requirements (e.g., Shahhosseini 2015), in food service in the form of letter grades of cleanliness posted on restaurants (e.g., Simon et al. 2005), and in education in the form of pay-for-performance and other school performance metrics (e.g., Bjorkman and Svensson 2009; Stecher et al. 2009). These systems rely on the so-called long path to accountability: being designed to empower citizens to respond—both through voice and exit (Hirschman 1970; 1976)—when the quality of public services (in our case public schools) deteriorates.

While previous work has explored whether citizens respond to performance signals attached to schools and other institutions, little work has specifically examined how the racialized elements of performance accountability systems affect the racial dynamics of schools, communities, and local elections. In particular, though policies like No Child Left Behind convey racially-tinged information signals (i.e. whether black students have failed, for example), virtually no work has precisely isolated the effect of these racial signals on whether and how citizens of various racial groups respond. That is, while we know quite a bit about whether citizens respond to overall school performance signals—which are a composite of several subgroup failures—we know much less about whether such aggregate effects are driven by the performance of salient racial in- and out-groups. Indeed, there are reasons to suspect that citizens from different racial backgrounds may respond differently when failure of a given school occurs

specifically because their racial subgroup is failing to perform at a high level, or, in contrast, a salient out-group. Further, while previous treatments have focused on citizens' voice-based responses to performance information (through examinations of voter turnout and incumbency reelection rates), little work has explored whether citizens also use exit—voting with their feet when they receive a signal that their school's performance has deteriorated. To our knowledge, no work has traced exit patterns of various racial groups to the racial components of performance accountability regimes. In this paper, we examine both voice and exit responses. Specifically, we explore how citizens respond to racial performance signals—examining, for example, how citizens of various races respond when they are told that their racial subgroup is the reason a given school is failing comparing this to how they respond when failure is attributable to salient out-groups.

To do so, we unpack the effects of No Child Left Behind's (NCLB) school-based racial failure signals. We leverage this seminal education policy and a unique dataset to estimate the causal effect of race-based school failure signals, or, put differently, the effect of schools' failing to make AYP *because of* low performance by various racial in- or out-groups. Our dataset provides one of the few combinations of large-scale school-level administrative records with large-scale individual-level voter administrative records (for the few exceptions see Holbein 2016 and Hastings et al. 2007). As an important innovation, we use as outcomes both validated measures of citizens' use of voice *and* of their use of exit. Our identification strategy leverages a regression-discontinuity design (RDD) examining schools close to NCLB's racial group failure cutoffs. This approach avoids endogeneity problems found in observational studies, thus allowing us to draw causal inferences. At the same time, this approach also preserves a greater degree ecological validity over possible lab or survey experiments studying the effect of

racialized performance cues on voice and exit, given that it comes from a real-world policy application.

We find that whites and minorities react to race-based failure signals in substantively different ways. White citizens' responses do not appear to be identity-based in the sense that whites are responsive to failure signals (both through voice and exit) regardless of the source. Whites respond to both in- and out-group signals in their voice and exit decisions. In contrast, African Americans appear to be more race-sensitive: being most likely to mobilize and vote in board elections when failure is attributed to their in-group. Moreover, while Black citizens are responsive to out-group signals through exit, they are much less likely to exit in response to failure by students of their own racial group. Put differently, African Americans do not exit poor performing schools when it is their group that is struggling; instead, they rally together by mobilizing in greater numbers in school board elections that follow. Our evidence suggests that this lack of exit in response to school black students' failure is likely not completely explained by a lack of resources or exit options for African-Americans in these schools, but instead is more consistent with a group identity response.[1] In short, while Whites are responsive overall across multiple response domains and failure sources, African Americans' responses show more nuance—being consistent with an racial group consciousness based explanation of performance responsiveness.

Our results show that racial cues embedded in performance accountability regimes can

[1] We readily admit that we are not fully able to conclude why we see these differences and that our data only suggests a particular motivation. One of the tradeoffs of using real-world, administrative data is that it limits our ability to pursue the specific attitudes driving these response patterns. While we find clear differences in responsiveness to racial information, our findings also highlight the need to better understand why race influences how individuals process and react to performance information (a point we return to in the conclusion).

affect individuals' retrospective responses, both through voice and exit. They suggest that citizen behaviors are not just the result of pure information content—as previous performance accountability studies have implied—but of an interaction of information with core group memberships and identities. Our findings also imply that the ever-increasing number of performance-based education policies, like NCLB, have differential effects on citizens of different racial groups. Perhaps most troubling, these results show that while performance information promotes accountability at the ballot box it also distorts the racial makeup of schools—in certain instances increasing white flight from low performing schools, promoting already troubling levels of racial segregation in public schools (Ladd, Clotfelter, and Holbein 2017). Our results find broader meaning in a context where citizens increasingly receive information from the government in a number of public and private domains, including on those involving the relative performance of various racial groups in shared social environments.

Background and Conceptual Framework

Previous work has emphasized that citizens' responses to performance information depend heavily on their ability to receive and act upon such signals, the salience of signals, and citizens' underlying levels of knowledge (Grant and Keohane 2005; Lassen 2015). Whether citizens are theoretically likely to respond to performance information has been hotly debated. On the one hand, if citizens were previously uninformed about the performance of public institutions, they may be likely to respond to this new information (Pande 2011). Similarly, if voters are partially informed about public performance, performance information may act as an alarm that alerts them to the deterioration of public performance (Holbein 2016). However, if actors are already fully informed when the information is sent (or simply miss or ignore the performance signal), there may be little to no public response (Ahn and Vigdor 2014).

High quality causal studies have tended to show that citizens do, indeed, respond to performance information from performance accountability regimes. In the work most similar to ours, Holbein (2016) shows that citizens respond to NCLB's overall school performance signals. Using administrative data from North Carolina and a regression discontinuity design, he finds that citizens respond to overall school failure both by increasing their level of voter turnout in school board elections and by exiting when schools marginally fail to make AYP cutoffs. Our work departs from and builds on Holbein (2016) in important ways. Though using data from a similar source and a similar identification strategy, our work uses a different set of treatments to answer a fundamentally important (unanswered) question—focusing on the effect of racialized information component of the overall signal of school failure. In so doing, our approach answers substantively important lingering questions in the performance accountability literature related to the performance of vitally important racial subgroups, racial identities, racial contexts, and racial differences in exit patterns and voter turnout propensities.

Other performance accountability studies have mostly found that school performance signals shape citizen evaluations (Charbonneau and Van Ryzin 2012; Chingos et al. 20012; Jacobson et al. 2013; Kogan, Lavertu, and Peskowitz 2016b) and prompt ballot box responsiveness on the part of citizens both in the US (Berry and Howell 2007; Barrows 2016; Clinton and Grissom 2015; Payson 2017; Rich and Jennings 2015) and in other countries (e.g. Toral 2016).[2] These results are also consistent with what scholars generally find in non-school settings. For instance, an abundant literature examines the effect of other types of performance accountability regimes in developing countries. This research also tends to show that citizens

[2] For other examples of citizens reacting to school performance information see also Scherer (2012) and Clinton and Grissom (2015).

respond to signals of the performance of local elected officials (Banerjee et al. 2011; Boyne et al. 2009; Chong et al. 2015; Ferraz and Finan 2007; Grossman and Hanlon 2014). This research parallels performance studies from developed settings. Indeed, the causal evidence overwhelmingly suggests that citizens respond to performance information, leading some to conclude, "that voter behavior is malleable and that information about the political process and politician performance improves electoral accountability" (Pande 2011, 215).[3]

While each of these studies is important in their own right, our work offers two important extensions. First, our work is one of the few studies to explicitly examine an alternate type of citizen response: exit. Most of the studies mentioned above examine whether citizens respond to negative performance shocks by voicing their dissatisfaction at the ballot box—by increasing their voter turnout in local elections, voting out incumbents, or being less likely to support school funding referenda. However, almost no work explores the possibility that citizens simply leave when they get bad news about their schools' performance. This is unfortunate given exit's prominent role in reforms like NCLB and in contemporary education policy more generally. Indeed, when signing NCLB into law, President Bush said, "there must be a moment in which parents can say, I've had enough of this school. Parents must be given real options in the face of failure in order to make sure reform is meaningful" (Bush 2002). Beyond even policy design, exit has long played a role in theoretical treatments of performance and accountability (Hirschman 1970; Hirschman 1976; Warren 2011; Vallier Forthcoming). Our approach examines the relatively unexplored possibility that citizens react to performance information by transferring from failing schools or by moving.

[3] Not all studies have shown that citizens respond to performance information (e.g. Olken 2007; Kogan, Lavertu, and Peskowitz 2016a).

In its simplest form, the idea is that citizens have two possible responses when they become concerned or dissatisfied with governments' (or other groups') performance: they can use *voice* (communicate a complaint or a proposal for change); or, they can *exit* (physically or cognitively leave or transfer to another public provider). Given the connected nature of these two behaviors, theorists have long noted the importance of studying these in tandem; yet, at present, few works have done so, with a few important exceptions (e.g., Holbein 2016; Rich and Jennings 2015; Hastings and Weinstein 2008).[4] This gap is unfortunate as studies that examine voice or exit in isolation may make misguided conclusions about performance policies' effectiveness if citizens use one channel, but not the other. This threat may be especially salient in local contexts: where opportunities for *both* voice and exit are abundant. The absence of studies that explore exit responses to school performance information parallels a broader gap in the literature on performance accountability regimes more generally, which tends to ignore, or only mention in passing, the possibility for citizens to vote with their feet.[5] Given the prominence of exit in theoretical accountability models and in policy design, this is an important gap in the literature.

Second, our work is different from previous work (including that closest to ours; Holbein (2016)) in that we examine a different treatment. Rather than exploring aggregated performance signals, as other work has predominantly done, our approach is to explicitly unpack how individuals of different racial groups respond when in- and out-subgroups fail, and as a result, cause schools to fail. This approach is vitally important given the salience of race and the racial makeup of schools to school enrollment decisions. Indeed, previous research in this area has

[4] While each of these three studies looks at exit, all fail to explore the racial dynamics of exit in response to race-based failure signals.

[5] For other examples of empirical examinations of exit in other contexts see Hastings, Kane, and Staiger (2008).

provided clear evidence that parents make enrollment choices based on the racial composition of schools, independent of students' performance on standardized tests (Schneider and Buckley 2002; Bifulco and Ladd 2007; Ladd, Clotfelter, and Holbein 2017). To our knowledge, only one other paper has attempted to isolate the effect of the racialized component of school performance signals. Flavin and Hartney (2017) explore the relationship between racial group performance and school board elections in California. They find that voters appear to respond to the performance of White students, but not Black and Hispanic students. Their paper is different from ours in that it does not focus on the responses of various racial groups, but rather the electorate as a whole; does not estimate the effect of exogenous performance shocks (i.e. failure or passage), but rather focuses on the observational relationship between continuous performance and local election outcomes; and does not consider exit as a possible response type. With this notable exception, most of the research on performance accountability has ignored racial dynamics. Insofar as racial and ethnic identities have been discussed in the context of previous treatments of performance-based accountability, they have been mentioned as a problem for which performance information is the solution. For instance, in developing contexts it is thought that performance information can help citizens overcome intransigent habits toward ethnic voting (Banerjee and Pande 2011). Still, rather than being an explicit part of examinations of performance accountability, race and ethnicity is only addressed tangentially. This is unfortunate as No Child Left Behind's race-based performance signals were billed as a key part of this seminal education reform. Yet, up until present the race-based dynamics of citizens' responses to racially-tinged performance information has been largely lumped together with the other components of student performance information.

Race and Information Responsiveness

There are a number of reasons why we might expect that race would affect responsiveness to performance information. Most of what we know about race and information responsiveness comes from political science research on the roots of political attitudes and behaviors. For instance, there continues to be large variation in participation rates across different racial groups in a multitude of political behaviors (Leighley and Nagler 2013; Verba, Schlozman, and Brady 1995; Wolfinger and Rosenstone 1980).[6]

In addition to the differences in resources afforded these two groups, minority social and group orientations provide an additional source of motivation beyond the traditional cost-benefit analysis of voting as traditionally understood in political science. Scholars have recently argued that participation in politics (and especially among minority groups) is a form of social expression (Garcia Bedolla and Michelson 2012; Gerber, Green, and Larimer 2008; Han 2014; Bobo and Gilliam 1990; Gay 2001; Segura and Woods 2007; Washington 2006). Group consciousness is a group identity that has been politicized through perceptions of the group's disadvantaged social standing and beliefs that collective action is the most effective way to better that situation (Brown 1931; Dawson 1995; Ferguson 1939; Jackman and Jackman 1973; Gurin et al.1980; Matthews and Prothro 1966; Miller et al. 1981). Group consciousness shapes political behavior by encouraging group members to consider group motivations and become more politically involved than they would otherwise (Dawson 1995; Miller et al. 1981; Olsen 1970; Shingles 1981; Verba and Nie 1972). Because of the systemic inequalities minority groups face (Dawson 1995; McClain et al.2009), minorities, and especially African-Americans, have a much higher sense of shared fate or linked fate than whites (Sanchez and Vargas 2016) which leads to

[6] While the disparity between whites and non-whites continues to exist, this gap has shrunk considerably in the last few decades (Leighley and Nagler 2013).

higher group solidarity (Hoston 2009) and provides a shield against threats through an emotional link to others within the group (Brown 1931; Shelby 2005). These attachments then also influence political engagement (Dawson 1995; Valdez 2011; Valenzuela and Michaelson 2016) and responses to information about how in-groups should act politically (Garcia Bedolla and Michelson 2012). However, few studies have examined how information responsiveness to racialized cues varies by racial subgroups (for an exception, see Garcia Bedolla 2005), specifically, and those that do have focused on how that information shapes the opinions of minorities rather than their actions and behaviors (Abrajano and Lundgren 2014; Masuoka and Junn 2013).[7] Moreover, while there is a burgeoning literature on how racial identity shapes attitudes and voice-related behaviors, we simply do not know how identity shapes citizens' decision to use exit—largely because there is such scant research on exit responses overall, much less those crossed with race.

Limitations in the previous literature most directly related to the topic at hand aside, from previous work we might expect that racially-tinged performance cues would shape behaviors across racial groups differently. After all, research has shown that racial minorities are more likely to have a sense of linked fate, which would cause them to more closely associate their well-being with the well-being of others in their group and shapes political behavior (see McClain et al. 2009 for a full review).

Theoretical Expectations

[7] Additionally, while there is a long literature on the influence of black racial threat on white behavior, there is little on minority reactions to white majorities and even those studies that do consider minority reactions to out-groups focus primarily on whites with minority reactions only addressed in passing (e.g. Glaser 2003, but for an exception see Masuoka and Junn 2013)

Based on previous research on race, political behavior, and performance accountability, we can derive a set of expectations about who, when, and how individuals are likely to respond to in- and out-group failure signals. We outline our set of expectations in Table 1 below, which shows what we might reasonably expect theoretically given previous work. We emphasize that these predictions are nondeterministic; that is, there are sensible reasons to suspect that these predictions might be wrong. We outline these theoretical predictions not to say they are guaranteed, but rather to give the reader a sense of how previous work frames our expectations and to give structure to our multiple analyses below.

We start with the responses of African Americans and then turn to those of whites. Based on work showing that linked fate is higher among African Americans (Dawson 1995; Miller et al. 1981; Olsen 1970; Shingles 1981; Verba and Nie 1972), ex-ante we suspect that a shared minority identity and higher perceptions of linked fate among minorities generally makes it more likely that Blacks will be more likely to use their voice and mobilize at the ballot box in response to Black student failure (hence our prediction in cell [6] in Table 1). This sense of linked fate helps Blacks overcome a lack of resources that are so vitally important for engagement in the political realm (Leighley and Nagler 2013; Verba, Schlozman, and Brady 1995; Wolfinger and Rosenstone 1980). Linked fate is so named because of an individual's association of their own well-being with the well-being of others in their group (Dawson 1995; Matthews and Prothro 1966; McClain et al. 2009). Individuals who have a high level of linked fate characterize harm done to any group member as a form of harm to the whole group (Brown 1931; Dawson 1995). Under this framework, in-group signals should trigger a response by driving in-group participants to mobilize as a means of trying help a shared group that is struggling to meet established performance thresholds.

Simultaneously, differential levels of linked fate might lead us to expect a different response for black citizens in the exit domain (cell [8]). Here, linked fate may keep African Americans from exiting when their group fails. For a group with strong social connections and a belief that the well-being of others in their own group affects them as well, exiting may be seen as an especially acute form of abandonment and detrimental to their own well-being (Dawson 1995; Hoston 2009; McClain et al. 2009). Hence, we expect that African Americans unlikely to vote with their feet and leave their school and their peers as a result of Black student failure.

For African Americans, responses to out-groups are ambiguous. In the voice domain (cell [2]), African Americans may respond to white failure if they see failure among this out-group as somehow impacting the well-being of their own group. On the other hand, however, Blacks may fail to respond if they use group cues as a heuristic for the overall performance of the school. Given that these competing forces, we leave this prediction ambiguous. Similarly, in the exit domain, African Americans may respond to out-group signals by exiting given the lack of linked fate failure constraints. Here again, a white failure signal may set off alarm bells about the performance school, which are not counteracted by group identities that nudge Blacks to stay in the school—hence leading to more Black students exiting. However, on the other hand, Blacks' willingness to exit overall may be constrained by other factors such as resources or choice options. Hence, Blacks' out-group responses are unclear; the responses are fascinating, but could theoretically go either way.

For whites, the predictions based on previous literature are conditioned on a slightly different set of factors. While whites have a less refined sense of linked fate (Sanchez and Vargas 2016), they may have greater access to a set of relevant individual resources, which the resource model or political participation predicts would enhance responsiveness overall

(Leighley and Nagler 2013; Verba, Schlozman, and Brady 1995; Wolfinger and Rosenstone 1980). Indeed, previous performance accountability work has shown some evidence that more affluent citizens are more likely to respond to overall school failure signals (Holbein 2016). Hence, though they lack a sense of linked fate, whites may respond to in-group signals by increasing their levels of political engagement (cell [1]).

Lacking a sense of linked fate also has implications for the exit domain; that is, whites may be more likely to exit public schools when their group fails (cell [3]). For whites, exit may not seem like group abandonment, but rather a response necessitated by circumstances. Enhanced resources and attentiveness among whites may mean that this group may be more likely to respond to out-group signals. If whites have higher attentiveness to performance signals overall, they may be more likely to respond to out-group signals. Here, out-group signals would serve as a proxy that something is wrong overall; that voice and exit are viable options because performance has deteriorated.

Table 1: Expectations for In-Group and Out-Group Responses in the Voice and Exit Domain

	White Voice	Black Voice	White Exit	Black Exit
White Failure	[1] Increase (greater attentiveness/resources)	[2] Ambiguous	[3] Increase (greater attentiveness/resources)	[4] Ambiguous
Black Failure	[5] Increase (greater attentiveness/resources)	[6] Increase (linked fate)	[7] Increase (greater attentiveness/resources)	[8] No effect (linked fate)

In sum, we have reason to suspect that race affects responses to racialized performance information. What remains unclear, however, is whether the prior research about candidate race and political behavior in other areas that inform our expectations transfer to performance accountability systems. Even more lacking is our understanding of whether information and race interact to play a role across citizens' use of a prominent alternative to voice—exit. Ultimately the nature of performance responses is an open empirical question.

Empirical Case: No Child Left Behind

To study the responses of racial groups to the racial information embedded in performance accountability systems, we use data from the education policy No Child Left Behind (NCLB). NCLB was signed into law on January 8, 2002 as a bipartisan reform. This reform evaluated schools' performance on yearly student learning outcomes and publicized performance information of individual schools and the performance of racial subgroups within those schools.

By signaling low performance, reformers hoped to mobilize parents, and specifically minority parents, to use voice to express their dissatisfaction or to simply leave when schools deteriorated. In short, NCLB was passed as a means of enabling students to seek a better education by helping them hold school leaders accountable and providing them with school choice when otherwise locked in failing public schools. Though NCLB was rewritten in late 2015 (as a part of the Every Student Succeeds Act), many elements of the policy remain intact. NCLB is still widely considered by many to be the "most far-reaching education polic[ies] … over the last four decades" (Dee and Jacob 2010, 149).[8] By many accounts, its passage fundamentally altered how public schools in the U.S. operated—placing greater emphasis on measurement and publicizing of performance.

Under NCLB, schools sent performance signals to their surrounding communities with the intention of empowering local communities to act, both through voice or exit, when school performance declined.[9] These signals consisted of discrete pass/fail signals for the overall school performance and for the performance of racial subgroups.[10] This information was disseminated

[8] The Every Student Succeeds Act continues to require that schools publish performance information but provides schools with much more flexibility in doing so, while also attaching fewer consequences to low performance.
[9] This is made clear in the text of NCLB, which repeatedly mentions its intention to increase "participation ... in school planning, review, and improvement" (ESEA 2002, 115 STAT. 1456).
[10] This information is presented with a spreadsheet that provides the percent proficient and the pass/failure status for the overall school and racial subgroups.

through formal letters from local school officials, an official website that received substantial web traffic (especially after a school failed), and various other informal channels (Holbein 2016).[11] Residents in the surrounding community had strong reason to pay attention to the overall performance of their school and that of racial subgroups within their schools (which partially determined overall school performance, as we explain below) regardless of whether they had children enrolled in schools, as housing values vary discontinuously at school performance cutoffs (Black 1999; Figlio and Lucas 2004).[12]

NCLB's naturally occurring race-based signals embedded in the law provide us with a unique opportunity to examine *if* citizens respond to negative performance signals about their in- and out-groups. This setup also allows us to explore *how* citizens respond to group signals varies; either by using voice to express their displeasure to elected officials and/or by exiting failing schools.

Data

To explore the effect of these embedded racial performance signals, we use a unique combination of data that draws from school administrative records and voter records. This data source combines information from the North Carolina public school system with the validated voting behavior of registered voters in that state in the public-use voter file. Because North

[11] NCLB states "the state educational agency shall publish, and disseminate to parents and the public, information on any corrective action ... through such means as the Internet, the media, and public agencies" (Section 1116(c)(10)(E), ESEA). Decentralized channels further amplify these messages, with local media outlets, parent networks, and realtors spreading school performance signals (Berry and Howell 2007; Black 1999).

[12] While NCLB does also signal Hispanic (and likewise Asian, Multi-Race, and American Indian) failure, the sample size of schools reporting performance is small given the laws providing exemption of schools with fewer than 40 individuals in a given subgroup (the number of exemptions given is continuous at the school failure cutoff). As such, any RDD analysis would be severely underpowered for Hispanic (and other groups') failure.

Carolina has long collected student-level information for all public-school students, we leverage this unique dataset to document the impact of race-based school failure signals both on voter turnout in school board elections and on exit from public schools.[13]

In this analysis, our primary independent variable is whether racial subgroups failed to make NCLB's student subgroup adequate yearly progress (AYP) thresholds causing the school to fail to make AYP overall.[14] The failure determination is described in detail in the methods section below, as it has direct bearing on the identification strategy used. Generally speaking, however, subgroup failure is determined by low student subgroup performance on standardized tests and overall failure is determined by subgroup failure. NCLB creates the mechanisms by which low performing schools signal to their surrounding communities that they have failed, while higher performing schools do not.

Our outcomes of interest are two-fold. First, as our measure of citizen voice we examine voter turnout—a standard measure used in accountability studies (e.g Banerjee et al. 2010; Chen 2013).[15] In this paper, we restrict the measure of turnout to elections where a school board race is on the ballot: where school performance is most relevant. In North Carolina, these elections generally occur during May in even-year primaries, with some exceptions.[16] The matching of

[13] Having the rich data available in the state comes at the cost of perhaps limiting generalizability. Using North Carolina as the test case mitigates some of these concerns, as it is the tenth most populous state in the United States, and has a diverse population living in both rural and urban settings and voter turnout that is comparable to overall national rates.
[14] This is drawn from North Carolina school performance data from 2003-2011.
[15] Other commonly-use proxies for voice focus on how citizens vote in relevant elections. Unfortunately, commonly used metrics such as vote share for the incumbent are not available in North Carolina, as the elections data for school board races does not include incumbency status.
[16] School board elections occur every other year. To simplify our analyses, school performance signals are matched to the voter data in the next school board election; put differently, in our analysis we do not use performance information in off-years for the main effects estimation. This is because there is not a theoretically-compelling or empirically straightforward way to incorporate this information into our single cutoff RDD framework. That said, we do use off-

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 education 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, especially historically. Thus, to fit the two data sources together, we follow the approach outlined by Holbein (2016). This approach matches citizens to the school that minimizes the Euclidean distance (as the bird flies) between home addresses and public schools.[17] In this paper, we use a match based on the closest elementary school. The elementary school match is done to preserve statistical power; after all, as in all states, in the North Carolina sample there are considerably more elementary schools (≈1300) than middle (≈600) or high schools (≈500). If a high or middle school match were to be used, we would effectively cut our statistical power by more than half (because this would drastically reduce the number of treatment clusters).[18]

This approach provides us with the distinct advantage of matching all of the

year information when we look at treatment effects by previous failure status (see Table A.6 and Figure A.5). 70.7% of districts (53.4% of schools) have their school board elections in May. In practice, this means that there are some schools where a differential amount of time has passed between when the signal was released and when the election is held (see Table A.4). In robustness checks, we test whether this matters for the signals' efficacy in the Online Appendix (see Figure A.7 in the Online Appendix). We use turnout data from board elections held 2004-2012.

[17] We intentionally do not force this match to put individuals in their correct school district. Doing so is difficult given the same issue we face with schools (a lack of historical boundary maps). But, we turn this potential weakness into a strength; after all, not forcing a district match allows us a post-matching benchmark of the quality of the match (see Table A.5 in the Online Appendix).

[18] As in Holbein (2016) who shows that overall school failure signals' effects do not vary by the school level matched to, our results remain similar, though less well powered, when we use the middle and high school matches. Power is more of a concern here than in Holbein (2016) as subgroup failure is sometimes missing given NCLB's small subgroups exemption. Holbein (2016) does not face this issue because he chooses over the available subgroups. We are constrained to the subgroups individually given our interest in racial signals.

approximately 7 million registered voters in North Carolina to their closest school. Moreover, robustness checks of a sample of matches to actual school boundary maps indicate that geographic matching approximates assignment matching sufficiently when comparing schools near NCLB's failure cutoffs.[19] This makes it very unlikely that the geographic matching procedure biases the regression discontinuity estimates outlined below. In essence, while this match does sometimes classify citizens in a non-assigned school, this mismatch only introduces additional noise into our estimates and attenuates them towards zero (Wooldridge 2009, 316-322).[20] Both of these forces act to make finding any effect more difficult—in this way, it serves to make our conclusions more conservative. Simply put, if we still find a significant effect, we can be more confident given this additional source of stochastic error (for more information on the school-voter match and checks using the limited available school boundary maps, see the Online Appendix).

Our second outcome measures citizen 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). Given that these measures are kept in-house by the NCERDC, the matching issues we face with the voter data are not a problem here—in this dataset, exit in a given school is cleanly and exactly matched to school performance in that same school. This large-scale administrative dataset documents the

[19] We were able to benchmark our geographic match to actual school assignment maps for a sample of schools during the 2009 school year. The school assignment maps came from the School Attendance Boundary Information System (SABINS). Among this sample, mismatch between our geographic approach and school assignment maps was balanced at the overall ($p=0.74$), white ($p=0.23$), and black school failure cutoffs ($p=0.18$) (see Figure A.4).

[20] In our application, mismatching could be conceptualized as measurement error in the independent or in the dependent variable. Given that we report several positive and statistically significant results below, it is appropriate to mark our results as being conservative: perhaps having inflated error variances and some bias towards zero.

flow of students into and out of the North Carolina public schools at the individual-level. These files do not count graduating or dropping out of school, but instead document the number of student transfers to another school in the district, transfers to a school outside the district, transfers out of the state, and transfers out of the public-school sector altogether (be it to private schools or homeschooling environment).[21] Given that some forms of exit—e.g. out of state moves—make some forms of voice less likely, as a robustness check we take these forms of exit out, which does little to our results (see Table A.9 in the Online Appendix). That is, most of the forms of exit we observe don't actually involve physically moving, but rather an exit from a given public partnership (a type of behavior that Hirschman (1970), and others, still call "exit"). As this data is contained in the NCERDC files, it is already linked to public school performance.[22]

Methods

To estimate the effect of the race-based failure signals embedded in NCLB on voter turnout and exit, we leverage discontinuities in school performance signals provided through NCLB using a regression discontinuity design (RDD). As is well established, observations that are sufficiently close to an arbitrary discontinuity are separated primarily by exogenous shocks (Butler and Butler 2006; Imbens and Lemieux 2008; Lee and Lemieux 2010). RDDs leverage this continuity or local randomization, to establish treatment and control groups that are similar on observables

[21] As with all administrative data, there is likely some misreporting in the number of student exits, with some schools under- and others over-reporting the number of exits. Some schools may even fail to report any student exits (as is the case with a small fraction of schools in our sample). Still, it is important to note that there is no incentive to misreport so under No Child Left Behind, as failure near the cutoff is not a function of exits.

[22] When voter turnout is the measure, we rely on the racial groups as identified in the North Carolina voter file. When exit is the outcome, we rely on the racial identity of the students documented in the school administrative data.

and unobservables—allowing us to be confident about the causal effect and avoiding some of the ecological validity problems associated with survey or lab experiments.[23]

In our application, treatment consists of a school-based subgroup failing to make adequate yearly progress (AYP), thus causing a school to fail and triggering the process of sending a signal to the surrounding community that the subgroup (and school) has failed. Control schools—those that marginally make AYP in a given subgroup—receive no such racialized signals. 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 many students fail to reach proficiency, the subgroup fails.[24] NCLB complicates this by providing two exemptions. The first exempts the subgroup if students within that group improve sufficiently from one year to the next (passing with growth or "safe harbor"). The second provides an exemption if the subgroup is sufficiently close to passing (passing with "confidence interval"). If either of these exemptions pulls a subgroup above the failure cutoff, that subgroup passes. Despite these complications, determining subgroup failure is simple because school performance is documented in the school administrative files—we know very clearly the subgroups that failed and those that passed in a given school.

[23] An alternate way to measure the causal impact would be to use survey or lab experiments that randomly assign one group to receive a negative signal. While causally robust, this approach limits ecological validity and may face ethical concerns given the possibility of negative identity signals having deleterious effects on citizen attitudes and behaviors (e.g., Arceneaux and Butler 2016). Our approach skirts these issues, while preserving internal validity.

[24] A less well-known rule requires schools to measure and report "other academic indicators" (or, OAIs for short), which in many states include attendance (in elementary schools) or graduation rates (in high schools). In practice, almost all (97.9% of schools among white students and 96.4% of schools among black students) schools pass on these OAIs, with most schools tightly bunched near the 100% threshold. As such, following previous work on No Child Left Behind (Ahn and Vigdor 2015; Holbein 2016; Holbein and Ladd 2017; Traczinsky and Fruehwirth 2014), we exclude these metrics in calculating our running variable scores and in our discussion below.

These exemptions, however, complicate the calculation of the running variable. Traditional regression discontinuity approaches use a single metric for the running variable. With multiple channels for passing in each subgroup under NCLB, the rule used for choosing the running variable must account for these. To do so, we use the procedure developed by Ahn and Vigdor (2014) specifically for the NCLB case (theirs mirrors the approach of Jacob and Lefgren (2004) and Matsudaira (2008) in other contexts). Their approach follows the intuition behind the codified rules of NCLB. This approach is done in two steps. The first chooses one channel (overall, growth, or interval) for each given subgroup. The decision rule used 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 students' performance would have to deteriorate for their subgroup to not pass through at least one channel.[25] The second step chooses one subject—either reading or math—to represent that subgroups' running variable score. The decision rule used to choose between these two subjects is to choose the one that has the *lowest* score. The intuition behind this decision rule is that, under NCLB's rules, if either subject causes a subgroup to fall below the threshold, that subgroup fails. Hence, the running variable is specified by the metric that is the furthest from the

[25] We use growth from one year to the next because NCLB incorporates it into the pass/fail determination. A small group of schools do very poorly one year, only to improve quite a bit the next. This type of dramatic improvement primarily occurs under the African American subgroups. As a result, the value of the running variable for these schools can be quite large—bounded only by the % improvement that they see above the improvement threshold (10%). These large outliers, however, play a relatively small role in our estimates because we use a triangle weight (which places greater emphasis on schools close to the cutoff) and because in the main estimates we set these outliers aside. The results do not change either if these are included or excluded.

cutoff, or the marker that would need to improve the most for the school to pass.[26]

Under this approach, 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 the school administrative 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 academic indicators used to determine school failure status.[27] Because we are not able to perfectly categorize schools with the proximity measure, fuzzy regression discontinuity design (RDD) is required (e.g., Ahn & Vigdor 2014; Matsudaira 2008; Holbein 2016).

Fuzzy RDD uses an instrumental variables approach to adjust for non-compliance (Angrist and Pischke 2008). Like in randomized-control experiments, Fuzzy RDD uses treatment assignment as an instrument for treatment receipt. In our application, treatment assignment is predicted subgroup failure (based on the running variable score) and treatment receipt is whether or not a subgroup actually failed to make AYP. The limited non-compliance in our application comes from schools that are marked failing when they should actually be passing, or vice-versa, given their performance. As non-compliance of this type is relatively rare, the instrument is sufficiently strong to satisfy the assumptions of instrumental variable models (Stock and Yogo 2005).[28]

Equation [1] shows a simplified form of our fuzzy RDD models. Each of the variables in the model is indexed at the racial group (g), school (s), or year (t) level.

[26] This assumes that ordered scores improve or decline in an order-preserving fashion.
[27] The exact interval used is not made public and some provisions for NCLB's other academic indicators are unclear. As such, we use the intervals employed by Ahn & Vigdor (2014).
[28] The standard F statistics for instrument strength are provided in the results tables below. The exact first-stage coefficients vary depending on the model and bandwidth run, but all are very large (in many cased being near 0.7-0.8, with 1 being the highest) and are all highly significant ($p<0.001$). This effectively scales up the size of our TOT estimates 20-30% relative to the ITT.

$$F_{sgt} = \gamma_0 + \gamma_1(P_{sgt}) + g(R_{sgt}) + X_{s-gt} + \xi_{sgt} \quad [1]$$

$$Y_{sgt} = \beta_0 + \beta_1(\hat{F}_{sgt}) + g(R_{sgt}) + X_{s-gt} + \varepsilon_{sgt}$$

In the first-stage, actual subgroup failure status (F_{sgt}) is estimated as a function of the subgroup's running variable score R_{sgt}. (F_{sgt} may alternatively be described as school failure *because* of a given racial subgroup given NCLB's provision of the whole school failing if one subgroup fails.) In this paper, we model the running variable with non-parametric, local-linear specification, denoted as $g(R_{sgt})$. This non-parametric approach allows us to use a data-driven process for modeling the running variable and, as such, avoids some of the shortcomings associated with parametric methods (Gelman and Imbens 2014). Also included in the first stage is the excluded instrument determined by the running variable (P_{sgt}). The simultaneously estimated second stage produces the causal effect of signaled failure on the outcomes of interest (Y_{sgt}): school board turnout and the number of exits from a given school. Our unit of analysis is the weighted school-year.[29] This approach preserves the rich nature of the data, while allowing us to deal with the clustered nature of the data and improve computation time.

In the Online Appendix, we show that our discontinuities are balanced across a number of observable characteristics (see Table A.1) and there is a lack of evidence of precise sorting (Figure A.1). As a result, in our models we simply control for the failure status of the non-examined racial subgroups (X_{s-gt}) to ensure that we are truly picking up on the effect of a given

[29] Models are weighted according to the number of registered voters (when turnout is the outcome) or students in the school (when exit is the outcome).

group signal and not shared treatment imbalances (Table A.2).[30,31] This approach makes comparisons of the effect of crossing a racial group failure threshold among schools that have the same failure status on other group performance dimensions. This ensures that our estimates are not picking up on the effect of other groups' failure. Instead, we are picking up the effect of interest: having a school fail because of the racial group we examine.[32] Hence, this allows us to estimate the effect of three cutoffs: overall failure (as a benchmark), white failure (accounting for black and other subgroup failures), and black failure (accounting for white and other subgroup failure).

Before we move to the results, a word about statistical power is in order. By design—given the bandwidths and weights used—RDD is a lower powered method for causal hypothesis testing (Lee and Lemieux 2010; Ahn and Vigdor 2014).[33] In our application, with millions of

[30] Unfortunately, our empirical case is not well situated to estimate a multi-cutoff regression discontinuity model—with subgroup discontinuities estimated simultaneously in the same model. The subgroups' running variable scores contains too much nonoverlapping missingness due to exemptions for subgroups with few students in a given school. While we can be certain that schools that are missing failure status did not fail, we cannot know how close they are to failing if they did have enough students. That said, subgroup missingness due to small N cells appears to occur randomly near the school failure cutoff (Holbein 2016, see Table A.4).

[31] In our data, 666 schools failed in the white student category. Of those, about a third were when black students had not failed. In our total sample, 4,781 schools failed in the black student category. (This higher number is consistent with previous research in the state that shows that the majority of failed categories from disadvantaged, minority students; Holbein 2016; Ahn and Vigdor 2014.) Among this group, about 80% failed when white students did not. In the data, 382 schools failed with both white and black students.

[32] This approach distinguishes White and Black failure from the failure of other racial groups. This approach allows us to tease apart the effect of specific types of racial group failure—our objective in this paper. Because of NCLB's provision that if one subgroup fails, the entire school fails, we cannot distinguish between failing in a racial group failure when the overall school fails and racial group failure when the overall school passes. We can, however, control for the subgroup category that pools all students together. When we do so, the results do not change.

[33] We do not use the approach outlined by Wing and Cook (2013) to generalize our estimates away from the cutoff because "identification of such effects requires stronger assumptions than those required for identification at the cutoff" (Angrist and Rokkanen 2015, 1331).

voters and students in our combined file, it may be tempting to argue that statistical precision is not much of a concern. However, this reaction is misguided. Since individuals are clustered within schools, our effective sample size is equivalent to the number of school-year observations in our data (the level at which treatment is determined). Adjusting for this clustered nature of the data, as we do by collapsing to the school-year level, reduces our statistical power drastically. As such, in our RDD models below we take the approach of using wider bandwidths for our estimates provided in our tables. This ensures that we will have enough statistical power to elicit effects, particularly those that examine our treatments' possible heterogeneities. That said, we are very sensitive to the potential criticism that, in so doing, we are introducing bias into our model results.[34] Thus, in our results section below we make sure to highlight the results' robustness (or lack thereof) across a wide range of bandwidths—from narrower than the so called optimal bandwidth to our estimates that use all of the data available. These robustness checks show that the conclusions that we draw are robust and not sensitive to potential sources of bias in the setting used.

Results

We begin by examining the effect of overall failure on voter turnout by racial group. As mentioned before, in our models we control for the failure status of the non-examined racial subgroups to ensure that we are truly picking up on the effect of a given group signal. As shown in the first row of Table 2, consistent with our expectations, both black (p=0.019) and white

[34] In some respects, the choice of the bandwidth is a second order decision relative to the specification of the running variable. If the functional form of the running variable has enough flexibility to approximate the underlying shape of the data along the running variable, the bandwidth matters little. This is because flexible approaches, such as the local non-parametric approach that we employ, use a fairly narrow binwidth in which to make approximations of the underlying data structure. In short, while bandwidths matter, they do not trump getting the running variable specification right—as we attempt to do with our data-driven approach.

citizens (p=0.066) show some evidence (with varying degrees of precision) of responding to overall school failure signals by turning out to vote in school board elections at higher levels. In the specification run, the white effect is slightly less precisely-estimated than the black effect. However, voter turnout in local school board elections for both blacks and whites increases by about 3 percentage points as a result of the school failing to meet overall AYP standards.

Breaking these effects apart by the race-based source of failure shows that the effects of in-group performance signals are similar for blacks and whites. Whites are quite responsive to white failure—with in-group failure increasing white school board turnout by 7.0 points, p=0.033—as are African Americans to the failures of their own group—with in-group failure increasing their turnout by 7.1 points, p=0.04. As we would expect, based on our theoretical predictions, when government provides information that performance has deteriorated within one's in-group, both white and African American voters respond by mobilizing in local school board elections.[35] The exact reasons for this response—be they linked fate or resources—are not easy to elicit; however, what is clear is the fact that in-group signals elicit in-group responses.

Table 2: NCLB Signals' Effect on Voter Turnout

	(1)	(2)

[35] Some might reasonably wonder why overall failure signals elicit a smaller effect than race-based signals when overall failure is a function of the performance of these two racial subgroups. The answer comes down to the effect of NCLB's other subgroups (e.g. Hispanic, Asian, American Indian, and Multi-race), which appear to have (noisily-estimated) zero effect on white and black performance. To us, this makes intuitive sense given the comparatively small numbers of students in these groups; simply put, whites and blacks have comparatively less attentiveness to their performance. The result of combining averaging over all subgroups is to water-down the larger salient in- and out-group signals. We do not include these estimates because our interest is primarily with whites and blacks, and, more importantly, because estimating the effect of failure among these subgroups is complicated by NCLB's small cells exemption, whereby schools with small populations of student subgroups (<40 students) do not need to report the percent of students that are proficient. Hence, while we can know whether these groups failed or not, we cannot calculate the running variable for these groups. Given that North Carolina has comparatively few individuals among these subgroups—and, hence, many schools missing proficiency numbers for these groups—these null estimates are severely underpowered.

	DV: White Turnout	DV: Black Turnout
Overall School Failure Signal	0.025 [-0.002, 0.052]	0.036 [0.006, 0.066]
White Failure	0.070 [0.006, 0.135]	-0.008 [-0.088, 0.071]
Black Failure	0.095 [0.027, 0.163]	0.071 [0.003, 0.139]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. In the racial signal models, controls include the corresponding racial failure statuses that are not being tested. In–group signals are highlighted.

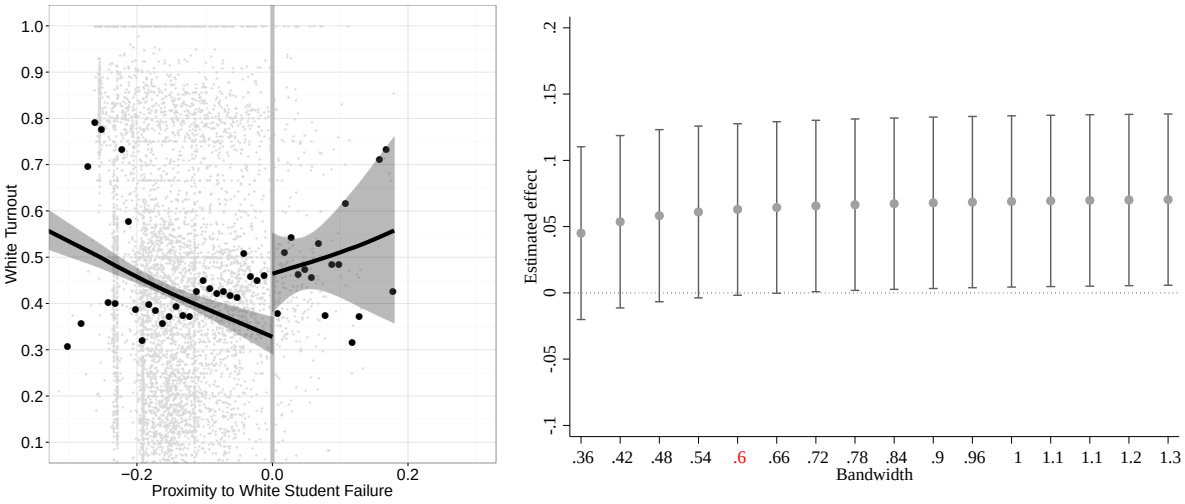
Responsiveness to out-group signals varies by race. White citizens seem to be more responsive overall, as we would expect, when black students are the reason a school failed—with these signals increasing turnout in board elections by 9.5 points, p=0.006. Importantly, this estimate is not statistically distinct from whites response to white failure; suggesting that white individuals are just as responsive to in-group signals as out-group signals. This result is consistent with white citizens simply being more sensitive to failure; rather than being driven in- and out-group identities, whites are simply concerned by failure, regardless of the source (a fact that is consistent with the exit results we show below). In short, when either black or white students fail, whites in the school zone are more likely to become involved in local school politics through the ballot booth. This suggests that in the voice domain whites' responses are not shaped by race; they simply are more responsive overall.

In contrast, African Americans do not appear to respond to out-group failure signals by mobilizing at the ballot box. When white students fail, there appears to be no discernible effect on the turnout of blacks in local elections. This result suggests that black citizens' retrospective evaluations in the voice domain are influenced by race. Our results are consistent with an explanation revolving around core identities fundamentally shaping whether minority groups

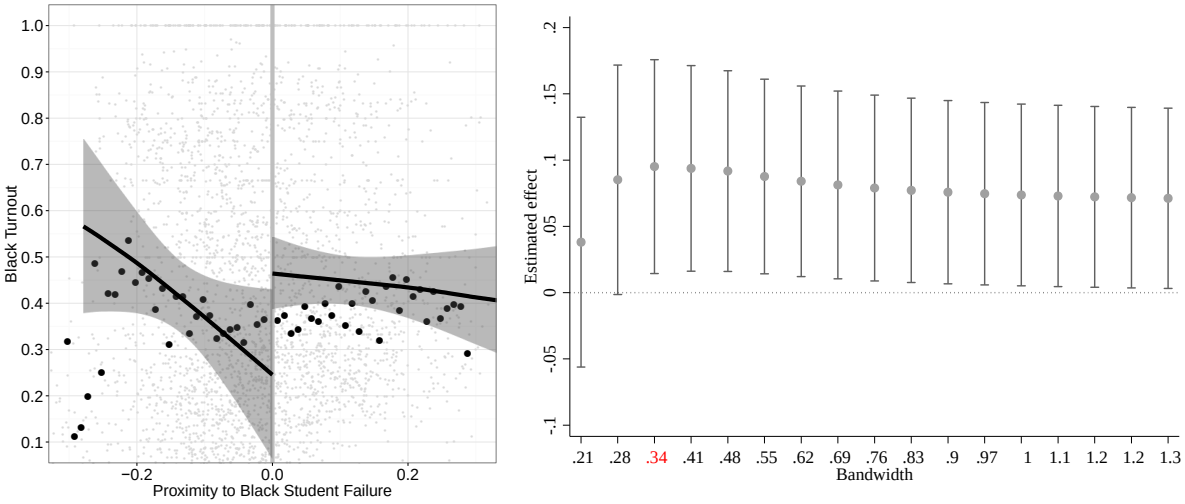
respond. Consistent with our expectations, we find that African-Americans are responsive to the failure of their own group. Moreover, a null out-group response is consistent with the ambiguous predictions we conveyed in Table 1. While we cannot explicitly test why this is, this finding fits within the theory of group interests that emphasizes a sense of linked fates among African Americans. As we explained previously, African Americans' heightened sense of linked fate could cause them to more closely associate their well-being with the well-being of others in their group and to react politically to threats to their own group rather than out-groups (Brown 1931; Dawson 1995; Hoston 2009; Matthews and Prothro 1966).

Figure 1: In-Group Signals' Effect on School Board Turnout

White Citizens



Black Citizens



Notes: Figure plots the effect of in-group signals on voter turnout in school board elections. The left panel plots the effect at the cutoff, with these estimates coming from the full bandwidth, a non-parametric local-linear specification of the running variable, school-level un-weighted data, and no controls. The right panel displays the effect estimates across bandwidths.

Our voice estimates shown in Table 2 are displayed graphically in Figure 1, which plots the vertical jump at the cut point along with the estimates across a range of bandwidths from the in-group failure signals (corresponding figures for the overall and out-group responses can be

found in the Online Appendix, see Figures A.2 and A.3).[36] As can be seen, there exists a discrete jump in local voter turnout at both of the in-group failure cutoffs. This effect is robust to the bandwidth chosen—with estimates being statistically indistinguishable across bandwidths. Estimates in narrower bandwidths are, as we would expect, less precise and underpowered. However, the effect estimates are not significantly different from one another across a range of bandwidths, suggesting that our results are robust and that narrower bandwidth results that do not quite achieve significance are the result of a lack of power. Given the low-powered nature of RDDs, the fact that the estimates are statistically indistinguishable across bandwidths and become significant when sufficiently powered is suggestive of a meaningful effect.[37]

Table 3 looks at the effect of race-based failure on school exit.[38] Whereas the effects of in-group student failure were consistent across racial groups for voter turnout, this is not the case for citizen exit. When white students are the source of school failure, an average of 87 more white students leave the school above and beyond those exiting marginally passing schools ($p<0.01$). This effect is moderate to large in size relative to the average school enrollment (roughly equivalent to 20% of the base rate) and is statistically distinct from zero and from the coefficient estimates of overall school failure. Simply put, whites are attentive to the failure of white students. Whites also show heightened levels of responsiveness to failure of out-groups, although with exit they are less sensitive to out-group signals than in-group signals. Signals of black student performance increases the number of white student departures by 37 compared to

[36] As can be seen in Figure 1, there are fewer schools that fail among white students than among blacks, and the schools that do fail among whites are closer to the cutoff. This should make intuitive sense given the well-known achievement gap between white and black students.
[37] While the black citizen response is smaller in the narrower bandwidth (0.21), it is not statistically different from estimates in slightly wider bandwidths.
[38] These effects are not entirely the product of the transfer sanction under NCLB. We show these effect estimates by sanction status in the online Appendix, see Figure A.5.

marginally passing schools ($p<0.03$); this response is more comparable to overall failure signal. In short, the results suggest that white individuals respond to both the failure of black and white students by voting with their feet, but slightly more to the performance of white students.[39] Previous accountability studies may have understated the effects of various performance regimes by ignoring exit as a potential viable alternate response to using voice.

Table 3: NCLB Signals' Effect on Citizen Exit

	(1) DV: White Exit	(2) DV: Black Exit
Overall School Failure Signal	25.00 [11.36, 38.64]	19.31 [2.84, 35.79]
White Failure	87.19 [54.15, 120.24]	106.57 [61.46, 151.68]
Black Failure	37.05 [4.05, 70.06]	-3.41 [-40.16, 33.34]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. In the racial signal models, controls include the corresponding racial failure statuses that are not being tested. In–group signals are highlighted.

In contrast, and again consistent with our expectations, we find that African Americans do not respond in the same way as whites to the failure of their own students. While more black students, on average, exit schools where white students marginally failed ($p<0.01$)—and when there is no sense of linked fate attached to the group who fails—there is no similar effect on exit for the marginal failure of black students ($p<0.86$). The effect estimates just discussed are robust across bandwidths—with the white response being present and the black response absent from very narrow to very wide bandwidths. In short, blacks do not respond to a failure signal from

[39] The effect of black failure on white exit is somewhat sensitive to the bandwidth that we use. For more information see Figure A2 in the Online Appendix.

their in-group by using exit, but do respond when their out-group fails.

Again, while we cannot explicitly test why this is, this finding also fits within the theory of group interests that emphasizes a sense of linked fate among African Americans. African Americans' heightened sense of linked fate could cause them to more closely associate their well-being with the well-being of others in their group and to be more hesitant to abandon others within their own racial group (Brown 1931; Dawson 1995; Hoston 2009; Jackman and Jackman 1973). These identities may be amplified when their group is the reason for a schools' failure; rather than trying to exit, black citizens stay in their current school (and simultaneously) are more likely to mobilize to vote (Table 2). However, when a sense of linked fate is absent (i.e. whites fail), blacks still exit and noticeably so. Both results are consistent with a stronger group identity among African Americans than whites.

To recap, whites respond regardless of racial cues: both through voting in school board elections and exiting failing schools. Their behavior in both domains seems to be driven very little by the source of failure; rather than being driven by group cues, they simply respond when failure information comes their way. In contrast, black citizens seem to be much more sensitive to the source of such failure signals—be they driven by in-group or out-group shocks. Blacks respond only to in-group signals via voice, and out-group signals via exit. When their group fails, they do not leave their school (like some whites), but they stay and mobilize (like other whites). These results suggest that white and black citizens are influence differently by racialized retrospective information. While race does not shape the retrospective evaluations of white citizens considerably, the same cannot be said for black citizens—whose retrospective behaviors seem to be fundamentally shaped by race.

Effect Heterogeneities

What drives the patterns we have shown in the last section? While identifying the direct effect of performance signals is relatively straightforward, teasing out causal mechanisms (as in all cases) is inherently difficult (Green, Ha, and Bullock 2010). In this section, we provide some descriptive evidence of how these effects vary. This allows us to shed some light on why citizens do or do not respond to racialized performance signals in their voice and exit decisions. Given space constraints, we cannot explore all theoretically interesting heterogeneities in the paper (for example, in the Online Appendix we consider how failure effects vary by school sanction status—see Table A.6 and Figure A.3—school district size—see Figure A.6—and timing of the signal—see Figure A.7).[40] However, here we focus on exploring whether racial context plays a role. We do so given the salience of the racial composition of schools in school enrollment choices and education policy more generally (Schneider and Buckley 2002; Bifulco and Ladd 2007; Ladd, Clotfelter, and Holbein 2017). We then turn to individual resources as a potential reason for the effects we observe.

To examine our results by racial context, we divide the sample into three groups based on the racial makeup of the school (the cutoffs we use for these are shown in the notes for these tables). So as to not bias our results towards finding statistical significance in one of the contextual groups, we split observations equally: with one third of the observations constituting the low racial composition, the middle third constituting the medium racial composition, and the top third constituting the high racial composition.[41] We then re-estimate the RDD models for

[40] We find some evidence that the effects are stronger in earlier years, consistent with an explanation signals schools sent may have been stronger. However, this pattern is not clear; only 4 of the coefficients in Tables 1 and 2 are in this direction and significant at the 5% level.
[41] If we divide the groups by percentage ranging from 0% to 33%, 33% to 67%, and 67% to 100% we find similar patterns of responsiveness. However, the smaller sample size in some of the groups affects statistical significance and skews towards those with arbitrarily large populations. For this reason, we deem the technique used in the paper preferable.

each subgroup. Because of space constraints, in the paper we display only some of these estimates. However, the other results are available in the Online Appendix.

Table 4: Voter Turnout Response by Racial Context (Black Citizens)

	(1) DV: Black Vote (Low Black)	(2) DV: Black Vote (Medium Black)	(3) DV: Black Vote (High Black)
Overall School Failure Signal	0.032 [-0.034, 0.098]	0.070 [0.017, 0.122]	0.029 [-0.012, 0.071]
White Failure	0.099 [-0.028, 0.227]	-0.017 [-0.134, 0.100]	0.061 [-0.102, 0.223]
Black Failure	0.091 [-0.32, 0.507]	0.079 [-0.033, 0.190]	0.066 [-0.019, 0.151]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. Racial groups are evenly divided across the three groups, with blacks in the minority representing schools with less than 14.9% in the school, racially mixed representing between 14.9% and 40.6%, and blacks in the majority above 40.6%. In the racial signal models, controls include the corresponding White, Hispanic, or Black failure statuses that are not being tested. In–group signals are highlighted.

Table 4 shows the effect of student failure on black turnout by the composition of African Americans in the school. The important differences here are across coefficients, rather than the statistical tests for differences from zero. As can be seen, we find no statistical differences across in-group signals (the last row, highlighted in grey) between black citizen reactions in low black population areas and high black population areas, with the point estimates and 95% confidence intervals overlapping.[42] Black citizens are just as responsive to the failure of black students in high black areas relative to low and medium black areas, with the coefficient sizes being statistically and substantively indistinguishable across the in-group signals (overall failure and out-group signals are also not statistically or substantively different across racial context). If

[42] We do see an effect in models without controls for other student failure. In these models, there is no effect of black failure on black turnout in areas with low and medium levels of black enrollment in the school, but there is a statistically significant effect in schools where blacks make up a larger percentage of the student population.

anything, the effects are largest in areas where African Americans are in the minority; but again, these differences are not significant at traditional levels. For African Americans in the voice domain, social networks appear to not play a role. Regardless of social context, in-group failure mobilizes African Americans. This result speaks to the power of the black failure signal under NCLB to the black community.

Table 5: Exit Response by Racial Context (Black Citizens)

	(1) DV: Black Exit (Low Black)	(2) DV: Black Exit (Medium Black)	(3) DV: Black Exit (High Black)
Overall School Failure Signal	0.96 [-3.81, 5.73]	22.68 [8.13, 37.23]	13.62 [-13.20, 40.44]
White Failure	10.70 [0.26, 21.14]	49.04 [25.19, 72.89]	155.75 [86.54, 224.96]
Black Failure	27.90 [7.10, 48.70]	33.01 [4.67, 61.34]	-20.40 [-71.56, 30.77]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. Racial groups are evenly divided across the three groups, with blacks in the minority representing schools with less than 14.9% in the school, racially mixed representing between 14.9% and 40.6%, and blacks in the majority above 40.6%. In the racial signal models, controls include the corresponding White, Hispanic, or Black failure statuses that are not being tested. In–group signals are highlighted.

Turning to the moderating role of racial context on the decision to exit schools, we find that racial context plays a role in the decision to exit failing schools, especially among minorities. Table 5 shows the effect of black student failure on the exit of black students across different racial contexts (with the last row focusing on in-group responses).[43] Our results show that the nulls mask an interesting heterogeneity; while blacks are no more or less likely to exit when they are attending schools where blacks make up a high percentage of the student body, they are more likely to respond to in-group failure by exiting schools where there are fewer black

[43] For white pa rents' exit response by white context, see the Online Appendix, Table A.3.

students. While in medium and low black areas, about 30 black students exit schools where blacks marginally failed, that effect disappears in schools where there is a large network of black students in the school. In these areas, instead of exiting, black families are more likely to keep their children enrolled in schools—perhaps rallying around students in their own in-group. Although we are not able to explicitly test if this is a result of group consciousness, this result is consistent with a group consciousness driven response; where networks are dense, exit does not occur among blacks in response to black failure, when the networks are sparser, exit does occur. This aligns with our expectations based on the literature on racial group empowerment, which has shown that group identities—like group consciousness—are strengthened by the presence of large numbers of others of that same racial group, especially among minorities (e.g., Bobo and Gilliam 1990; Barreto, Segura, and Woods 2004; Garcia Bedolla and Michelson 2012; Valenzuela and Michelson 2016; Spence and McClerking 2010). It seems that the same holds true for responsiveness to performance information with individuals' exit decisions. Blacks are especially unwilling to abandon their group when they are surrounded by members of their group, but are more willing to do so when they are not.

Interestingly, in dense social networks, blacks appear to be *more* likely to respond to white failure. It appears that the source of failure matters a great deal for the moderating effect of social context. When schools struggle and it is not one's ingroup that is causing failure, dense networks among black parents appear to encourage exit. The precise reasons for this difference are not clear and deserve attention in future research. They could be driven by an asymmetry in how black citizens leverage their social networks in schools. When their group fails, dense social networks prevent exit; but when another group fails, dense social networks encourage exit.

While fully exploring the roots of the patterns we observe is beyond the scope of this

paper, there are several possible explanations. Our evidence suggests that these key differences in exit patterns are unlikely solely the result of an inability of blacks to exit failing schools or because a lack of choice options for blacks in these areas. While we do not deny that access to resources disproportionately affect the mobility of minorities and whites, we find suggestive evidence that there are other forces that also impact the decision to exit failing schools that are different for whites and blacks. Specifically, there appears to be something about the size of the minority group itself that changes exit responsiveness: suggesting that ours are perhaps a product of linked racial identities.[44,45]

While a lack of resources is a reasonable explanation, we believe there are other explanations that may also impact these differences between whites and blacks. The most obvious is that while blacks do not exit high-density schools when blacks fail, they *can and do exit* high-density black schools when white students fail. When white students fail, black individuals in high black areas are *much more likely to exit* than blacks in low black areas. If these patterns were solely the product of differential resources or options to exit, we would not expect individuals in these areas to react by exiting the school at all: because they do not have the financial means or alternative schooling options that would allow them to do so. However, as noted in Table 5, in schools where blacks make up a sizeable portion of the student body the

[44] Aside from physically moving across school boundaries, students in North Carolina schools have many choices when it comes to where they can go when they exit a given school. The state has always been a leader when it comes to charter schools, with these being prominent in districts such as Durham, Charlotte-Mecklenburg, and Wake. The state also has a number of magnet schools. These alternate schooling options are especially prominent in the state's urban centers—where many of the African American students in the state are located.

[45] An alternative explanation may be that the schools are different in terms of the number of students enrolled, with higher density schools having higher populations. However, even if we look at exits as a percentage of the population in these schools, we still see similar results to what we document below.

failure of white students has a significantly and substantially *bigger* effect on exit than it does in schools where blacks make up a smaller portion. While 10–50 black students left schools where white students failed in low and medium black areas, on average 156 black students left schools where whites failed in high black areas (see the middle far right cell in Table 5). Thus, the size of the black population does not seem to inhibit or prevent blacks from exiting the school. Rather, it appears to be the racial dynamics at play that changes black exit behavior. While we do not deny that minorities struggle more than whites to move easily to other sectors of their community, we do find evidence that blacks do regularly change schools.[46] Second, when we formally incorporate proxies of individual resources into our models, the results do not change. The North Carolina education data contains the proportion of students on free or reduced price lunch—a good proxy for poverty rates in the school. When we include this measure as a formal control in our heterogeneity models—thus making the three groups equivalent in their underlying levels of resource, but varying their racial composition—the pattern of lower black responsiveness in high black areas remains the same. Net of the free-reduced lunch rate in the school, black individuals in high proportion black areas remain much less likely to exit when their subgroup fails than black citizens in less dense social networks—having about 40-50 fewer exits ($p < 0.03$).[47] Together, these two pieces of evidence suggest that even when taking into account individual

[46] In addition, in areas where there is a high percentage of minorities (and thus by extension areas that are more likely to be impoverished), whites continue to exit the system just as they do in areas where minorities make up a small portion of the system (and thus likely to not be impoverished) as seen in Table A3 in the Online Appendix. While whites in these areas may have more resources, combined, these two findings suggest that blacks are staying in the school zone not because they do not have the means to exit.

[47] These results come from a RDD model that specifies the running variable as flexibly linear, restricts the bandwidth to 0.2 units on either side of the cutoff, and includes black failure status, the proportion of black in the community, the interaction between the two, the running variable, an interaction between the running variable and failure status (to allow for flexibility on either side of the cutoff), and the proportion of free/reduced price lunch students in the school zone.

resources and exit options the differential effects remain.

Still, we admit that we cannot fully explain the underlying reasons for this action. These may be driven, in part, by parents' relative satisfaction with schools and the inputs which they weigh to be most important in their children's school, which tends to vary by racial subgroups (Friedman, Bobrowski, and Geraci 2006). While black families do have the ability to, and often do, transfer out of failing schools, they are much less likely to do so when members of their in-group surround them. One possible explanation that we outline above may be because of an enhanced sense of citizen loyalty in these densely homogenous social networks. Individuals in these networks may be more likely to have higher levels of group consciousness or linked fate and could be less likely to use exit as a response type because this might be seen as an abandonment of one's own group, and, as such, may come with negative psychological or social repercussions. However, regardless of the reasons driving these important responses, these effects are meaningful and show fundamental differences in how individuals of different racial groups respond to performance information.

Discussion

Performance information systems continue to grow as a policy-based means of increasing accountability: being prevalent now across education, health, crime, food, and development sectors. While an abundant, ever growing literature explores the effect of these performance systems, little work has been down to explore how these interact with the racial dynamics of schools, communities, and local elections. Utilizing the distinct racial signals sent by NCLB to surrounding communities about the success or failure of students of different racial groups, we find substantively meaningful differences in voice and exit among whites and non-whites.

While we are not able to explicitly test the exact reasons why we find these differences, they conform to expectations about how racial identity and group consciousness shapes the behavior African-Americans and Whites. Previous research has argued that racial information should contribute by "giving group members an understanding of their common interests" (Hardin 1995, 25) and by logical extension an understanding of how to act and participate politically (Dawson 1995). Shared circumstances and identities within a minority group shape behavior and encourage group members to become more involved when they perceive their group interests at stake (Dawson 1995; Miller et al. 1981; Olsen 1970; Verba and Nie 1972). Moreover, racial minorities are more likely to have a sense of linked fate which would cause them to more closely associate their well-being with the well-being of others in their group and to act in group interests rather than self-interests (Dawson 1995).

While whites respond to the failure of their own students by turning out to vote at higher rates and also by exiting failing schools, black individuals are, on the whole, much less likely to use exit as a response to their group's failure. While black citizens react by turning out to vote when their group fails, just as whites do, they are less likely to abandon fellow in-group students by exiting the school zone when their group is the reason for school failure. This aversion towards exit is especially evident in areas with higher black populations (which is often associated with higher levels of minority empowerment (Barreto et al. 2004)); with black citizens in high density areas being willing to pay the extra cost of not abandoning a school to support students of their own group. In short, these exit patterns suggest that these differences in racial responsiveness contribute to racial segregation in schools; indeed, this can be seen in aggregate enrollment numbers, with failure increasing black enrollments by 2-8 percentage points in the year after failure depending on the source of failure ($p<0.02$ in all cases). Our results suggest that

a lack of response by black parents is not (exclusively) the result of individual resource or choice constraints, but, instead, is directly related to the size of the minority group in the school. Although we cannot explicitly test the motivation why in this study, these findings suggest that group consciousness plays a role in the decision to act or react to new information. Future work would do well to unpack these effects further, perhaps in lab environments, paying specific attention to why groups do or do not respond to salient out groups (where the predictions are less clear). The findings for African Americans differential responses to white failure deserve additional attention specifically.

Our results have important policy implications. Speaking of this legacy-building legislation at the end of his presidency at a speech given at a school that was over 75 percent African-American, President George W. Bush explained that, "under this system, if your public school is failing, you'll have the option of transferring to another public school or charter school. And ... I view that as liberation. I view that as empowerment" (Bush 2009). While the goal of NCLB was to provide students—especially those disadvantaged minorities—with the ability to exit low performing schools, what we find is that the information that NCLB provides does not actually work in such a simple way. When considering the effects of performance accountability systems designed to elicit citizens' accountability-enhancing responses, it is important to remember that not all citizens are equally responsive. Racial minorities, in particular, are less likely than whites to leave their school community in the face of failure of students of their own racial group, and, as a result, failure of black students has no net influence on the exit of black students from a failing school.[48] This, perhaps in some ways, suggests that deepening segregation

[48] The downstream effects of these patterns on school performance and the behavior of elected officials are unclear. The question remains as to whether local officials responded to these differential patterns in citizen responsiveness, as they have shown to do so in other contexts (e.g.,

from white flight may also be explained through the flipped lens of many African Americans' behavior—which remain in their communities and mobilize in support of their students when schools fail, while whites, on the other hand, exit.

More generally, our results show that racial performance information shapes the way people process and react to political information. These suggest that whites and minorities react differently to race-based failure signals in a way that is consistent with theories of minority group identity. While previous work has shown that race is an important component of political action, we have demonstrated that race also matters in terms of how individuals respond to political information. This conclusion is evident in more traditional measures of citizen voice (i.e. turnout) and, also, through the less-explored channels of action including exit. This initial foray has delineated these differences and opened up avenues for new research about the effect of race on information processing and response to racial information. While our preliminary findings suggest these results are the product of racial identity and not resources, fully understanding the nature of race's impact on information responsiveness requires further research. Future work would do well to explicitly explore whether shared racial attitudes explain the patterns in voice and exit that we have documented here. While more remains to be done, our results suggest that who we are and the groups to which we belong conditions how we process and respond to information and, ultimately, how we behave across a number of domains.

Butler and Broockman 2011). Exploring this question, while valuable in its own right, is simply beyond the scope of this paper.

References

Abrajano, Marisa, and Lydia Lundgren. 2014. "How Watershed Immigration Policies Affect American Public Opinion Over a Lifetime" *International Migration Review* 49(1):70-105

Ahn, Tom, and Vigdor, Jacob L. 2014. "When Incentives Matter Too Much: Explaining Significant Responses to Irrelevant Information*" National Bureau of Economic Research Working Paper* #w20321.

Angrist, Joshua D., and Miikka Rokkanen. 2015."Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff." *Journal of the American Statistical Association* 110(512): 1331-1344.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Banerjee, A. V, Rukmini Banerji, and Esther Duflo. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2(1):1–30.

Barreto, Matt A., Gary M. Segura, and Nathan D. Woods. 2004. "The Mobilizing Effect of Majority-Minority Districts on Latino Turnout." *American Political Science Review* 98(1):65–75.

Barrows, Samuel. 2015. "Performance Information and Retrospective Voting: Evidence from a School Accountability Regime." PEPG Working Paper 15-03.

Bechtel, Michael M., and Jens Hainmueller. 2011. "How Lasting Is Voter Gratitude? An Analysis of the Short- and Long-Term Electoral Returns to Beneficial Policy." *American Journal of Political Science* 55(4): 852–68.

Bifulco, Robert, and Helen F. Ladd. 2007. "School Choice, Racial Segregation, and Test-Score Gaps: Evidence from North Carolina's Charter School Program." *Journal of Policy Analysis and Management* 26(1): 31-56.

Björkman, Martina and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124(2):735–69.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114(2):577–99.

Bobo, Lawrence and Franklin D. Jr. Gilliam. 1990. "Race, Sociopolitical Participation, and Black Empowerment." *American Political Science Review* 84(2):377–93.

Boyne, George A., Oliver James, Peter John, and Nicolai Petrovsky. 2009. "Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments." *Journal of Politics* 71(4): 1273-1284.

Brown, W.O. 1931. "The Nature of Race Consciousness." *Social Forces* 10(1):90-97.

Bush, George W. 2000. "Speech to the NAACP 91st Annual Convention." Baltimore, MD. 7/10.

---. 2002. "President Signs Landmark No Child Left Behind Education Bill." Hamilton, OH. 1/8.

--- W. 2009. "Speech at General Philip Kearny School" Philadelphia, PA. January 9.

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.

Charbonneau, Étienne, and Gregg G. Van Ryzin. 2012. "Performance Measures and Parental Satisfaction with New York City Schools." *American Review of Public Administration* 42(1):54-65.

Chen, Jowei. 2013. "Voter Partisanship and the Effect of Distributive Spending on Political Participation." *American Journal of Political Science* 57(1):200–217.

Clinton, Joshua D., and Jason A. Grissom. 2015. "Public Information, Public Learning and Public Opinion: Democratic Accountability in Education Policy." *Journal of Public Policy* 35(3): 355-385.
Dawson, Michael C. 1995. *Behind the Mule: Race and Class in African-American Politics.* Princeton: Princeton University Press.
Dee, Thomas S. and Brian A. Jacob. 2010. "The Impact of No Child Left Behind on Students, Teachers, and Schools." *Brookings Papers on Economic Activity* 2010(2):149–94.
Ferguson, Elizabeth A. 1938. "Race Consciousness among American Negros." *Journal of Negro Education* 7(1):32-40.
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.
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.
Flavin, Patrick, and Michael T. Hartney. 2017. "Racial Inequality in Democratic Accountability: Evidence from Retrospective Voting in Local Elections." *American Journal of Political Science* 61(3): 684-697.
Friedman, Barry A., Paula E. Bobrowski, and John Geraci. 2006. "Parents' School Satisfaction: Ethnic Similarities and Differences." *Journal of Educational Administration* 44(5): 471-486.
Garcia Bedolla, Lisa. 2005. *Fluid Borders: Latino Power, Identity, and Politics in Los Angeles*. Berkeley, CA: University of California Press.
Garcia Bedolla, Lisa and Melissa R. Michelson. 2012. *Mobilizing Inclusion: Transforming the Electorate through Get-Out-the-Vote Campaigns*. New Haven: Yale University Press.
Gay, Claudine. 2001. "The Effect of Black Congressional Representation on Political Participation." *American Political Science Review* 95(3): 589–602.
Glaser, James M. 2003. "Social Context and Inter-Group Political Attitudes: Experiments in Group Conflict Theory." *British Journal of Political Science* 33(4):607–20.
Grant, Ruth W., and Robert O. Keohane. 2005. "Accountability and Abuses of Power in World Politics." *American Political Science Review* 99 (1): 29–43.
Green, Donald P., Shang E. Ha, and John G. Bullock. 2010 "Enough Already About 'Black Box' Experiments: Studying Mediation is More Difficult than Most Scholars Suppose." *The Annals of the American Academy of Political and Social Science* 628(1): 200-208.
Grossman, Guy, and W. Walker Hanlon. 2014. "Do Better Monitoring Institutions Increase Leadership Quality in Community Organizations? Evidence from Uganda." *American Journal of Political Science* 58(3): 669-686.
Gurin, Patricia, Arthur R. Miller, and Gerald Gurin. 1980. "Stratum Identification and Consciousness." *Social Psychology Quarterly* 43(1):30-47.
Han, Hahrie C. 2014. *How Organizations Develop Activists: Civic Associations and Leadership in the 21st Century*. New York: Oxford University Press.
Hardin, Russell. 1995. "Self-Interest, Group Identity." Pp. 14–42 in *Nationalism and Rationality*, edited by A. Breton. New York: Cambridge University Press.
Hastings, Justine S., and Jeffrey M. Weinstein. 2008."Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *The Quarterly Journal of*

Economics 123(4): 1373-1414.
Heinrich, Carolyn J. 2009. "Third-party Governance under No Child Left Behind: Accountability and Performance Management Challenges." *Journal of Public Administration Research and Theory* 20(1): 59-80.
Heinrich, Carolyn J. 2002. "Outcomes–based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness." *Public Administration* R*eview* 62(6): 712-725.
Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.
Hirschman, Albert O. 1974. "'Exit, Voice, and Loyalty': Further Reflections and a Survey of Recent Contributions." *Social Science Information* 13(1):7–26.
Holbein, John B. 2016. "Left Behind? Citizen Responsiveness to Government Performance Information." *American Political Science Review* 110(2):353-386.
Holbein, John B., and Helen F. Ladd. 2017 "Accountability Pressure: Regression Discontinuity Estimates of How No Child Left Behind Influenced Student Behavior." *Economics of Education Review* 58: 55-67.
Hoston, William T. 2009. "Black Solidarity and Racial Context: An Exploration of the Role of Black Solidarity in U.S. Cities." *Journal of Black Studies* 39(5): 719–31.
Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142(2):615–35.
Jackman, Mary R., and Robert W. Jackman. 1973. "An Interpretation of the Relation Between Objective and Subjective Social Status." *American Sociological Review* 38(5):569-82.
Kinder, Donald R. and Cindy D. Kam. 2010. *Us Against Them: Ethnocentric Foundations of American Opinion*. Chicago: University of Chicago Press.
Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. 2016a."Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda." *American Journal of Political Science* 60(2): 418-435.
Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. 2016b. "Do School Report Cards Produce Accountability Through the Ballot Box?" *Journal of Policy Analysis and Management* 35(3): 639-661.
Kroll, Alexander, Milena I. Neshkova, and Sanjay K. Pandey. 2017. "Spillover Effects From Customer to Citizen Orientation: How Performance Management Reforms Can Foster Public Participation." *Administration & Society.*
Kroll, Alexander. 2017. "Can Performance Management Foster Social Equity? Stakeholder Power, Protective Institutions, and Minority Representation." *Public Administration* 95(1): 22-38.
Ladd, Helen F., Charles T. Clotfelter, and John B. Holbein. 2015. "The Growing Segmentation of the Charter School Sector in North Carolina." *National Bureau of Economic Research* (w21078).
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.
Leighley, Jan E. and Jonathan Nagler. 2013. *Who Votes Now? Demographics, Issues, Inequality, and Turnout in the United States*. Princeton, NJ: Princeton University Press.
Martin, Steve, Sandra Nutley, James Downe, and Clive Grace. 2016. "Analysing Performance

Assessment in Public Services: How Useful is the Concept of a Performance Regime?" *Public Administration* 94(1): 129-145.
Masuoka, Natalie and Jane Junn. 2013. *The Politics of Belonging: Race, Public Opinion, and Immigration*. Chicago: University of Chicago Press.
Matsudaira, Jordan D. 2008. "Mandatory Summer School and Student Achievement." *Journal of Econometrics* 142(2):829–50.
Matthews, Donald R., and James W. Prothro. 1966. *Negros and the New Southern Politics.* New York: Harcourt, Brace & World.
McClain, Paula D., Jessica D. Johnson Carew, Eugene Walton, Jr., and Candis Watts. 2009. "Group Membership, Group Identity, and Group Consciousness: Measures of Racial Identity in American Politics?" *Annual Review of Political Science* 12: 471-485.
Michener, Gregory, and Otavio Ritter. 2017. "Comparting Resistance to Open Data Performance Measurement: Public Education in Brazil and the UK." *Public Administration* 95(1): 4-21.
Mickelson, Roslyn Arlin, and Stephanie Southworth . 2005 "When Opting Out is not a Choice: Implications for NCLB's Transfer Option from Charlotte, North Carolina." *Equity & Excellence in Education* 38(3): 249-263.
Miller, Arthur H., Patricia Gurin, Gerald Gurin, and Okasana Malanchuk. 1981. "Group Consciousness and Political Participation." *American Journal of Political Science* 25(3):494–511.
Miller, Susan M. 2017. "How Electoral Incentives Shape Performance Measurement Efforts." *Public Administration* 95(3): 623-637.
Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, D.C: Georgetown University Press.
Moynihan, Donald P., and Patricia W. Ingraham. 2003. Look for the Silver Lining: When Performance Based Accountability Systems Work." *Journal of Public Administration Research and Theory* 13(4): 469-490.
Moynihan, Donald P., Sanjay K. Pandey, and Bradley E. Wright. 2011. "Setting the Table: How Transformational Leadership Fosters Performance Information Use." *Journal of Public Administration Research and Theory* 22(1): 143-164.
Nielsen, Poul A., and Donald P. Moynihan. 2017. "How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy." *Journal of Public Administration Research and Theory* 27(2): 269-283.
Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2).
Olsen, Marvin E. 1970. "Social and Political Participation of Blacks." *American Sociological Review* 35(4):682–97.
Pande, Rohini. 2011. "Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies." *Annual Review of Economics* 3(1): 215-237.
Payson, Julia A. 2017. "When Are Local Incumbents Held Accountable for Government Performance? Evidence from US School Districts." *Legislative Studies Quarterly* 42(3): 421-448.
Radin, Beryl A. 2000. "The Government Performance and Results Act and the Tradition of Federal Management Reform: Square Pegs in Round Holes?" *Journal of Public Administration Research and Theory* 10(1): 111-135.
Rich, Peter M., and Jennifer L. Jennings. 2015. "Choice, Information, and Constrained Options."

American Sociological Review 80(5): 1069-1098.
Sanchez, Gabriel R., and Edward D. Vargas. 2016. “Taking a Closer Look at Group Identity: The Link between Theory and Measurement of Group Consciousness and Linked Fate.” *Political Research Quarterly* 69(1): 160–74.
Schneider, Mark, and Jack Buckley. 2002. “What Do Parents Want from Schools? Evidence from the Internet.” *Educational Evaluation and Policy Analysis* 24(2): 133-144.
Shelby, Tommie. 2005. *We Who Are Dark: The Philosophical Foundations of Black Solidarity*. Cambridge, MA: Harvard University Press.

Shingles, Richard D. 1981. “Black Consciousness and Political Participation: The Missing Link.” *American Political Science Review* 75(1):76–91.
Simon, Paul A., Phillip Leslie, Grace Run, and Ginger Zhe Jin. 2005.”Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalizations in Los Angeles County.” *Journal of Environmental Health* 67(7): 32.
Spence, L. K. and H. McClerking. 2010. “Context, Black Empowerment, and African American Political Participation.” *American Politics Research* 38(5):909–30.
Stecher, Brian M. et al. 2009. *Toward a Culture of Consequences: Performance-Based Accountability Systems for Public Services*. Santa Monica, CA: RAND Corporation.
Stock, James H. and Motohiro Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” Pp. 80–108 in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D. W. K. Andrews, J. H. Stock, and T. J. Rothenberg. New York: Cambridge University Press.
Toral, Guillermo. 2016. "When are Local Governments and Bureaucrats Held Accountable for the Quality of Public Services? Evidence from Brazil's Education Sector." Working Paper available at SSRN #2776782.
Traczynski, Jeff and Fruehwirth, Jane C. 2014. “Spare the Rod? The Dynamic Effects of No Child Left Behind on Failing Schools.” *Working Paper.*
Valdez, Zulema. “Political Participation among Latinos in the United States: The Effect of Group Identity and Consciousness.” *Social Science Quarterly* 92(2): 466-482.
Valenzuela, Ali A., and Melissa R. Michelson. 2016. “Turnout, Status, and Identity: Mobilizing Latinos to Vote with Group Appeals.” *American Political Science Review* 110(4):615-630.
Vallier, Kevin. Forthcoming. "Exit, Voice, and Public Reason." *American Political Science Review*.
Verba, Sidney and Norman H. Nie. 1972. *Participation in America: Political Democracy and Social Equality*. New York: Harper & Row.
Warren, Mark E. 2011. “Voting with Your Feet: Exit-Based Empowerment in Democratic Theory.” *American Political Science Review* 105(4):683–701.
Wing, Coady, and Thomas D. Cook. “Strengthening the Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison.” *Journal of Policy Analysis and Management* 32(4): 853-877.
Wooldrige, Jeffrey M. 2009. *Introductory Econometrics*. Cengage Learning, 4e.

Online Appendix

When Your Group Fails: The Effect of Race-Based Performance Signals on Citizen Voice and Exit

[Not intended for publication in print versions]

I. Specification Checks

In this section, we examine four specification checks—first, the McCrary (2008) density test for precise sorting of the running variable; second, checks for pre-treatment covariate balance; third, tests for simultaneous failure of other racial subgroups; and, finally, the estimates provided in Tables 1 and 2 across various bandwidths of the running variable.

Ia. McCrary Density Test

Here we perform the McCrary (2008) density test for precise sorting. This test is recommended as a standard check of regression discontinuity models, though it has well-known limitations relative to other RDD specification checks (i.e. covariate balance). Indeed, as McCrary notes, "a running variable with a continuous density is neither necessary nor sufficient for identification except under auxiliary assumptions" (2008, 701). Simply put, when discontinuities in the distribution of observations arise, the McCrary density check cannot distinguish between whether these are the result of precise sorting or because the underlying performance distribution takes that shape, perhaps for institutionalized reasons (as we show below).

We are not the first to examine the distribution of AYP performance using a regression discontinuity framework. Some studies that have done so have shown continuity in the distribution of schools. For example, using data from Maryland, Hemelt shows that the performance distribution is smooth (2011, see Figure 6). Similarly, using data from Wisconsin, Chakrabarti (2014) shows that the performance distribution is smooth (see Figure 3). In other states, however, there is some evidence from other states of an imbalance close to the cutoff. As noted by some (Ahn and Vigdor 2015; Holbein 2016; Holbein and Ladd 2017; Traczinsky and Fruehwirth 2014), in the NCLB case a small cluster of schools may exist just above the just passing side of the failure cutoff. However, these studies have shown that this cluster is unlikely to be explained by precise sorting by schools. Rather, it is a byproduct of the safe-harbor (SH) and confidence interval (CI) exceptions that NCLB provides to schools as alternate ways of passing.

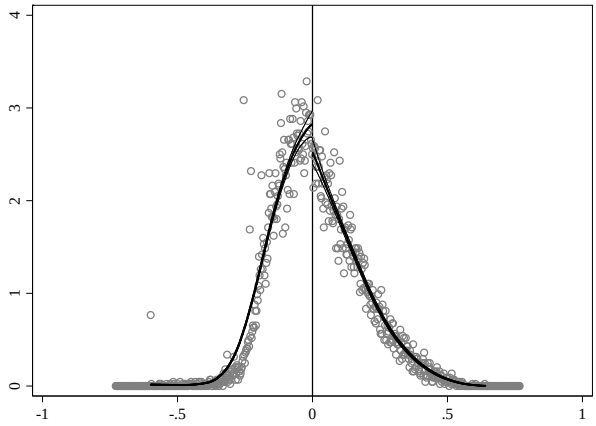
Indeed, when we remove these exceptions from the running variable calculation (as we show in our McCrary figures below), our running variable for overall school performance shows signs of balance around the cutoff. This suggests what appears to be, at first glance, precise sorting is actually a bi-product of the nature of AYP itself. This should make intuitive sense, as the SH and CI exceptions are built to push marginal schools over the cutoff (as we see in the McCrary distribution). Hence, it seems that AYP is, by design, built to produce a distribution that fails the McCrary check. Furthermore, in the discontinuities that we focus on in the paper—that for Black and White students—the distribution appears to be continuous both with and without the exceptions. This suggests that imbalances are the result of other, perhaps less salient, student subgroups.

Again, given the inherent limitations of the McCrary check, the covariate balance check is more informative of local randomization and the absence of precise sorting. Evidence supporting a lack of these conditions can be found in Table A.1 below, which demonstrates that schools are not clustered according to observable characteristics around our cutoffs.

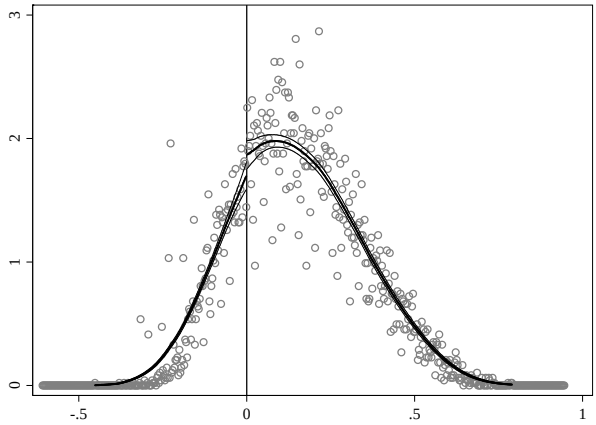
Figure A.1: McCrary Density Check

Overall Performance

With SH/CI Exceptions

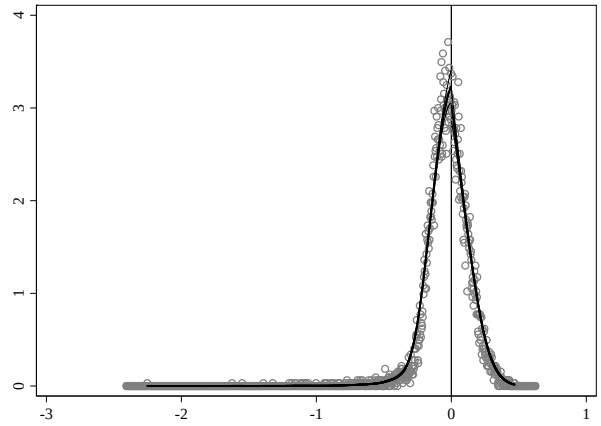


No SH/CI Exceptions

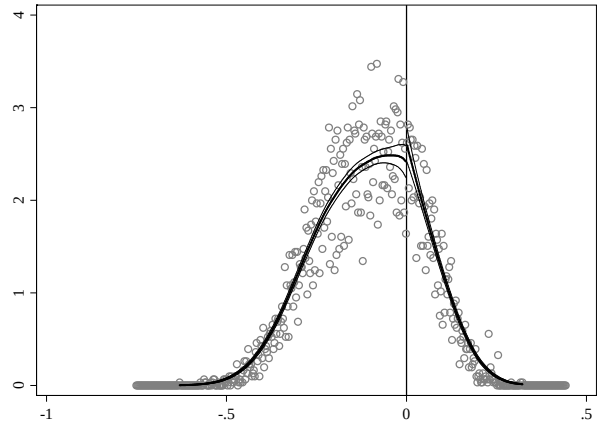


Black Performance

With SH/CI Exceptions

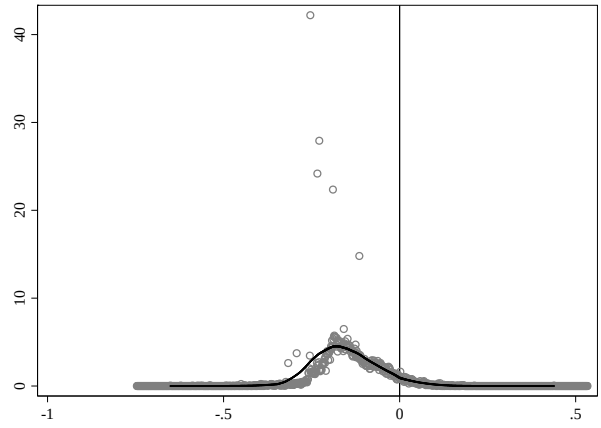


No SH/CI Exceptions

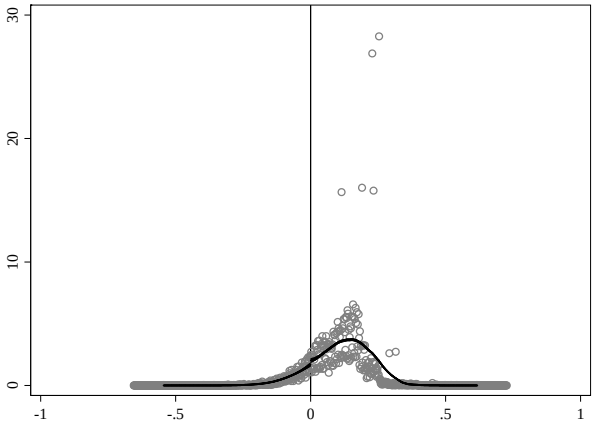


White Performance

With SH/CI Exceptions



No SH/CI Exceptions



Ib. Covariate Balance

Table A.1 shows the covariate balance across our cutoffs. As shown by others who have studied the AYP cutoffs (Ahn and Vigdor 2014; Holbein 2016; Holbein and Ladd 2017; Traczinsky and Fruehwirth 2014), we confirm that observable covariates are highly balanced at the cutoff. Of the 88 statistical tests provided in Table A.1, only 7 (7.9%) are statistically significant at the 5% level—only slightly more than what we would expect given the multiple hypotheses being tested here. Further, of the 7 covariates that are imbalanced, none of these occur in more than one of the four models run in the paper, most of the differences are relatively small substantively, and even fewer clear the Bonferoni or Sidak thresholds for multiple comparisons. Finally, of the lagged dependent variables—arguably the most important covariates for balance to get unbiased estimates—only 1 of 16 (6.3%) is imbalanced (but again not at the Bonferoni or Sidak thresholds).

Table A.1 Covariate Balance

Variable	Overall Failure (White Voters) β	p	Overall Failure (Black Voters) β	p	White Failure (White Voters) β	p	Black Failure (Black Voters) β	p
White Turnout	0.01	0.66	-0.02	0.54	0.02	0.71	-0.01	0.94
Black Turnout	-0.02	0.61	0.00	0.91	0.20	0.03	0.07	0.45
White Exit	-0.09	0.13	-0.01	0.14	-0.14	0.08	0.00	0.32
Black Exit	0.00	0.69	0.00	0.55	0.00	0.32	-0.01	0.32
Average Age of Registered Voters	1.04	0.06	1.28	0.14	-1.18	0.52	-0.58	0.74
Prop Registered Female	0.00	0.70	0.00	0.48	-0.01	0.22	0.00	0.94
Prop Registered Republican	-0.02	0.23	0.01	0.54	-0.08	0.04	0.04	0.23
Prop Vote in Person	0.00	0.95	0.03	0.09	-0.04	0.21	0.00	0.95
# Registered Voters	-77.44	0.26	-23.71	0.00	-37.45	0.06	-17.48	0.43
# Black Registered Voters	1.22	0.85	-14.33	0.12	-11.84	0.47	-21.95	0.26
Hispanic Turnout	-0.03	0.55	-0.05	0.45	0.01	0.96	-0.06	0.69
Prop Students Immigrant	0.00	0.62	0.00	0.14	0.00	0.24	0.00	0.21
Prop Students Homeless	0.00	0.87	0.00	0.97	0.00	0.73	0.00	0.06
Prop Students Parent College	0.03	0.40	-0.01	0.76	0.16	0.00	0.21	0.40
Prop Free/Reduced Lunch	0.00	0.89	0.06	0.00	0.00	0.88	0.06	0.22
Prop Students Limited English	0.01	0.22	0.01	0.27	0.02	0.13	-0.01	0.51
# Black Students	8.96	0.15	19.56	0.04	6.18	0.62	-2.02	0.9
# Hispanic Students Exiting	-0.01	0.20	0.00	0.15	-0.02	0.12	0.00	0.32
# Total Students Exiting	-0.10	0.13	0.00	0.99	-0.15	0.08	-0.01	0.32
# Students	-1.21	0.88	-114.09	0.00	-276.95	0.13	-131.89	0.11
Highest Grade Offered in School	-0.01	0.77	-0.04	0.38	0.03	0.72	-0.04	0.70
Earliest Grade Offered in School	-0.05	0.43	-0.08	0.36	-0.04	0.90	0.07	0.68

Note: Check for covariate balance at the various pass (C) / fail (T) discontinuities used in this paper. The second column shows the difference between treatment (failing) and control (passing) groups on these baseline covariates. Figures derived from a regression discontinuity model measuring the effect of school failure on these lagged outcomes. Like in our other models, the full bandwidth of schools is used. Frequency weights for the number in the respective voter groups. Running variable is modeled as a quartic polynomial. Models are clustered to the school-year level. The racial failure treatments mirror the models used in the paper--controlling for failure other racial subgroups.

Ic. Bias from Other Simultaneous Racial Group Failure

As we are examining multiple discontinuities, it is possible that one racial group's failure is confounded by another group's simultaneous failure. If this were the case, our estimates of white and black failure might be biased—picking up instead failure of other groups. Table A.2 tests this possibility, looking for failure across other racial groups at the white and black cutoff.

Table A.2 shows that schools in which the white students are close to the cutoff, but are exogenously nudged onto the failing side are discontinuously more likely to also fail among African American students. Hence, to account for this in our RDD models, we control for African American failure status in all of our models estimating the effect of White failure (we include the other racial subgroups as a precaution as well). Table A.2 shows a similar imbalance around the cutoff for Black student failure. Schools that are close, but arbitrarily fall on the failing side of the cutoff, have a higher likelihood of also failing with Hispanic students. Hence, similar to our white failure models, we include a control for Hispanic failure (again, we include the other racial subgroups as a precaution as well).[49]

Table A.2: Other Racial Subgroups Failure at the Cutoffs of Interest

	(1) DV: White Failure	(2) DV: Black Failure	(3) DV: Hispanic Failure	(4) DV: Asian Failure	(5) DV: American Indian Failure
White Failure	.	0.19 [0.01, 0.37]	0.01 [-0.13, 0.15]	0.02 [-0.02, 0.07]	0.00 [-0.01, 0.01]
Black Failure	0.01 [-0.02, 0.03]	.	0.15 [0.07, 0.23]	0.00 [-0.01, 0.01]	0.00 [-0.01, 0.01]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. Unfortunately, there are too few schools that have enough multi-race students to report failure status. Given this missingness, we cannot estimate the models for this subgroup. However, because this type of failure is so rare (occurring in less than 1% of school-years), it is unlikely to bias our estimates in the paper.

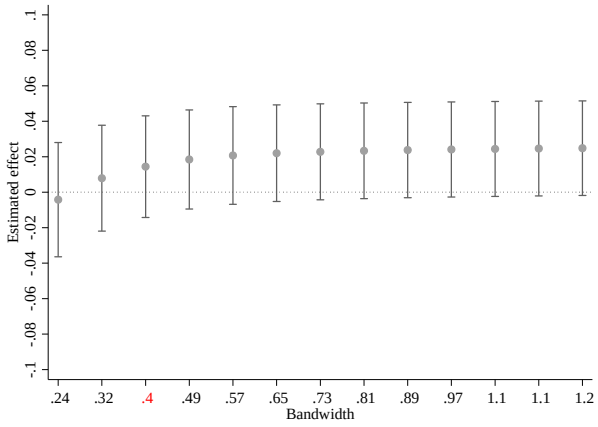
[49] We do not include failure on economically disadvantaged, students with disability, or limited English proficiency as these overlap multiple racial subgroups and thus suffer from multicollinearity with our measures of interest.

Id. Estimates Across Bandwidths

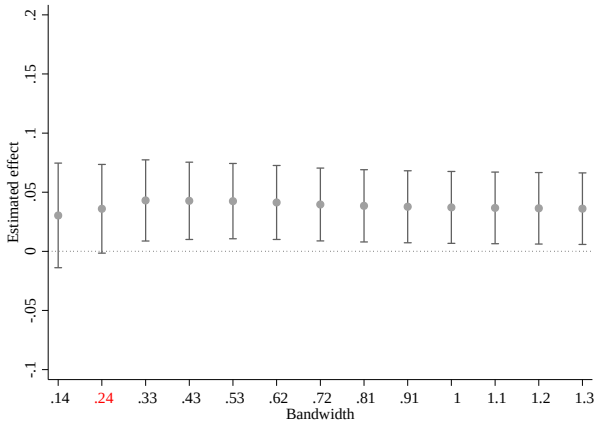
Figure A.2 Overall Turnout Estimates (Table 2) by Bandwidth

Overall

Overall Failure, White Turnout

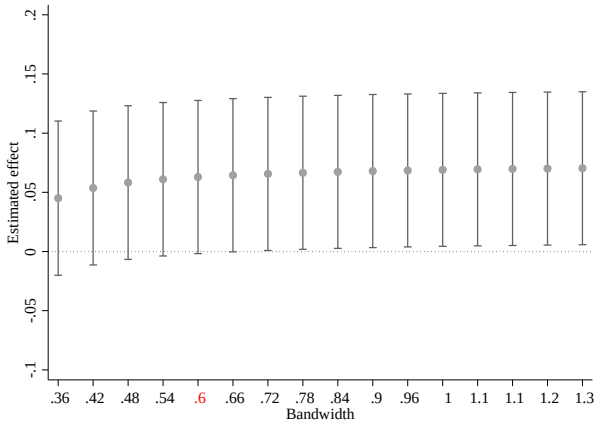


Overall Failure, Black Turnout

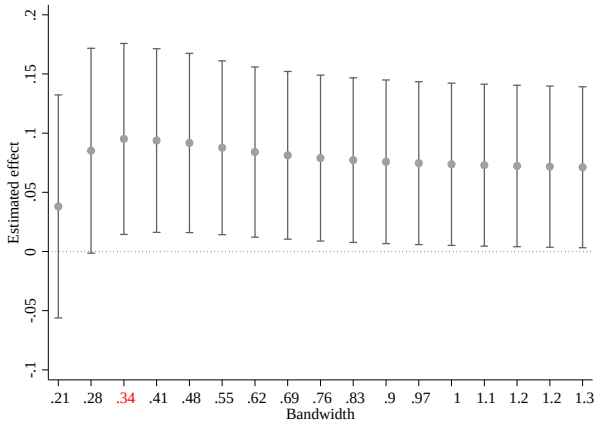


In-Group

White Failure, White Turnout

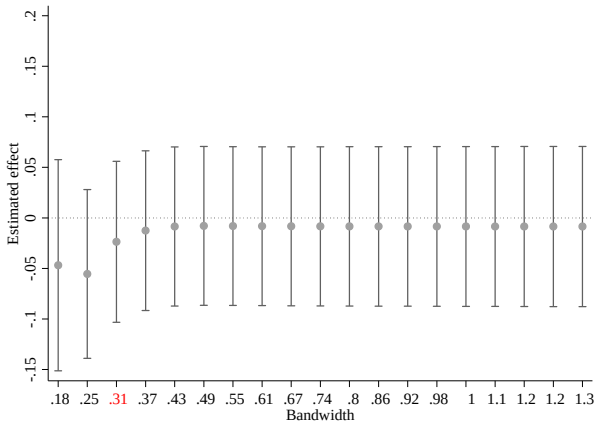


Black Failure, Black Turnout



Out-Group

White Failure, Black Turnout



Black Failure, White Turnout

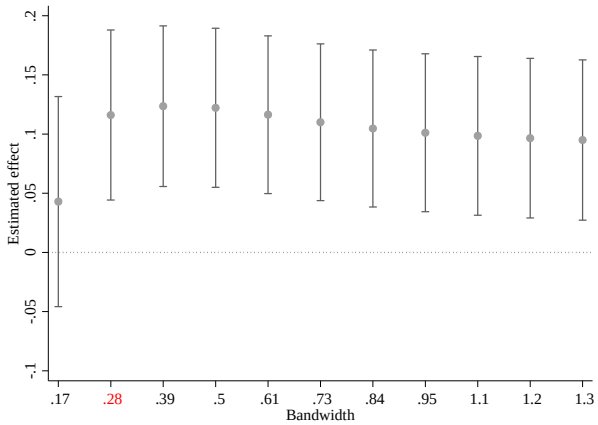
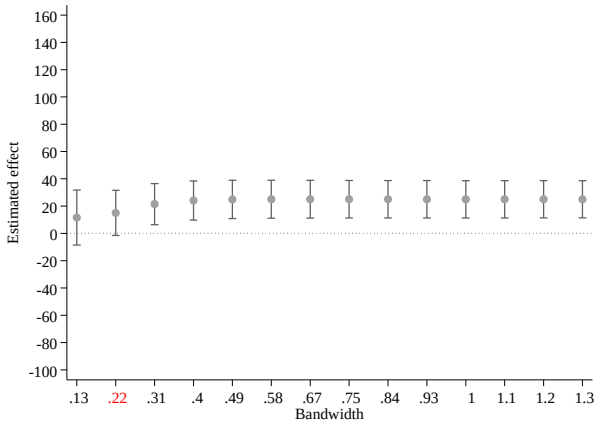


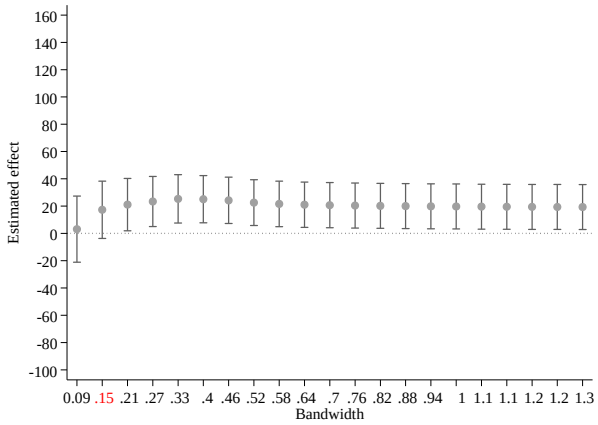
Figure A.3 Overall Exit Estimates (Table 3) by Bandwidth

Overall

Overall Failure, White Exit

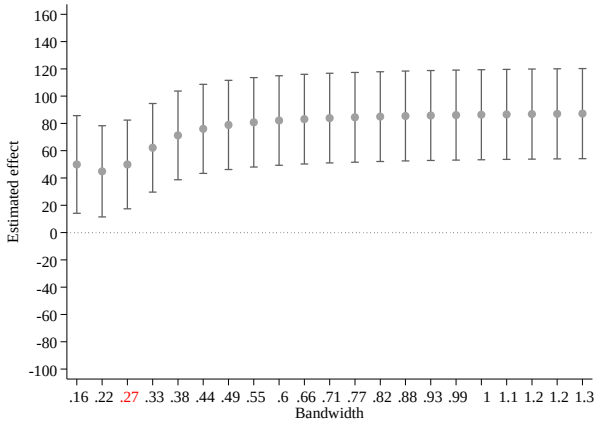


Overall Failure, Black Exit

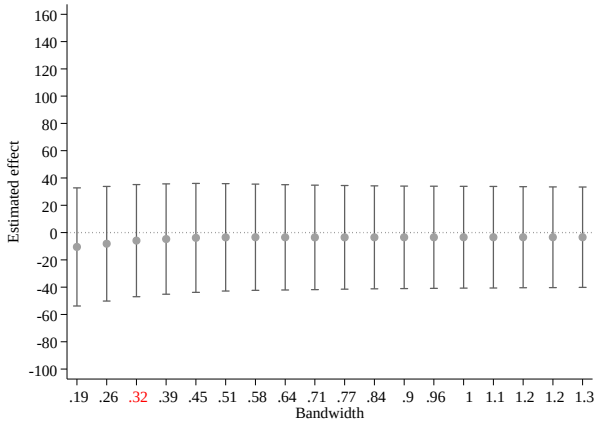


In-Group

White Failure, White Exit

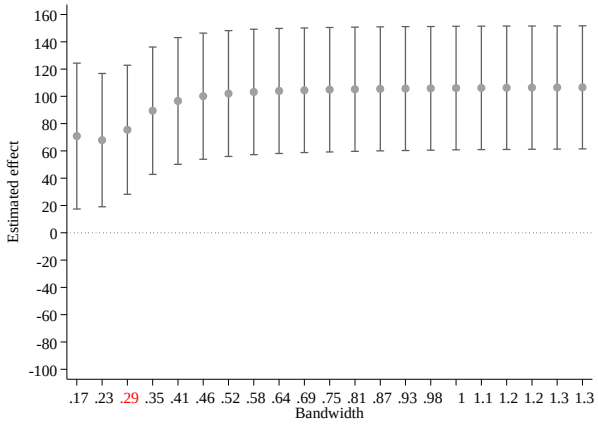


Black Failure, Black Exit

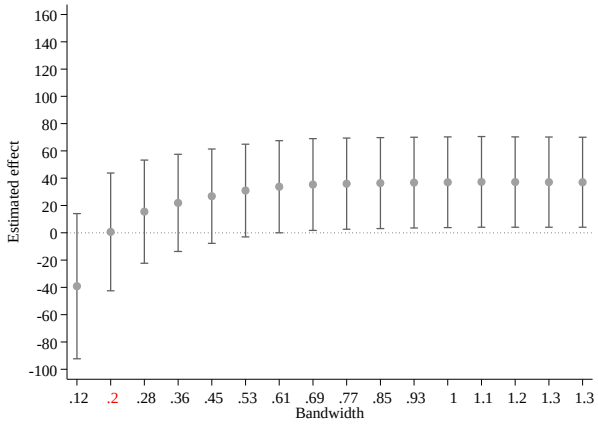


Out-Group

White Failure, Black Exit



Black Failure, White Exit



II. White Exit in High Density Minority Neighborhoods

One explanation to the lack of exit by black students from schools where blacks fail is that it is the lack of resources or exit options. We are hesitant to push that theory for a couple of reasons. First, as outlined in the text, we find that blacks are actually *more likely* to exit high density minority schools when whites fail than they are in low density minority schools. If blacks are exiting schools with high minority populations at higher rates than they are schools with low minority populations under the same circumstances, this suggests that it is not resources or lack of options.

In addition, in the text we show that when we formally incorporate proxies of individual resources into our models, the results do not change. The North Carolina education data contains the proportion of students on free or reduced price lunch—a good proxy for poverty rates in the school. When we include this measure as a formal control in our heterogeneity models, the pattern of lower black responsiveness in high black areas remains the same. Net of the poverty rate in the school, black individuals in high proportion black areas remain much less likely to exit when their subgroup fails than black citizens in less dense social networks—having about 40-50 fewer exits ($p < 0.03$).

Lastly, Table A.3 shows the effect of student failure on the exit of white citizens by racial context. In areas where there is a high percentage of minorities (and thus by extension areas that are more likely to be impoverished), whites continue to exit the system just as they do in areas where minorities make up a small portion of the system (and thus likely to not be impoverished).[50] Taken together, these two findings suggest that blacks are staying in schools not because they do not have the means to exit.

[50] Because the results presented in Table A.3 are divided evenly to ensure our findings are not biased by statistical power, the percentage of white students in the low white context does not exactly match the percentage of white students in the high black environment in Table 5 in the text. However, replicating the models to make the low white population match the high black population bracket produces virtually identical results.

Table A.3: Exit Response by Context (White Citizens)

	(1) DV: White Exit (Low White)	(2) DV: White Exit (Medium White)	(3) DV: White Exit (High White)
Overall School Failure Signal	14.21	22.53	26.32
	[0.53, 27.88]	[0.92, 44.14]	[4.66, 47.97]
White Failure	43.39	70.14	105.81
	[11.47, 75.31]	[39.58, 100.69]	[47.22, 164.39]
Black Failure	8.92	22.38	76.94
	[-23.60, 41.44]	[-18.87, 63.65]	[-10.16, 164.05]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. Racial groups are evenly divided across the three groups, with white in the minority representing schools with less than 42.0% in the school, racially mixed representing between 42.0% and 72.2%, and whites in the majority above 72.2%. In the racial signal models, controls include the corresponding white, Hispanic, or black failure statuses that are not being tested. In–group signals are highlighted.

Table A3 also can be used to interpret how whites respond to school failure signals. Table A3 shows the effect of student failure on the exit of white citizens by racial context. Across all contexts, whites appear to be willing to exit schools in response to overall failure and, even more-so, when signals are attached to their in-group.[51] The response to in-group signals appears to increase with the proportion of individuals in the in-group. In contrast to black parents, whites are actually more likely to exit their schools in response to in-group failure in areas where whites are a larger percentage of the student population. While white student failure causes about 43 white students to exit the school in low white schools, this effect grows to 106 students (more than twice as many) when looking at schools where whites are the predominant majority. In contrast to patterns in Black responsiveness, white exit grows in response to white failure as the percentage of white students in the school grows.

[51] Because the results presented in the paper are divided evenly to ensure our findings are not biased by statistical power, the percentage of white students in the low white context does not exactly match the percentage of white students in the high black environment the paper. However, replicating the models to make the low white population match the high black population bracket produces virtually identical results.

III. More Details on the Schools-Voter Match

Table A.4 shows how school performance was linked to voter turnout from school board elections. North Carolina school districts are required to hold school board elections in even-year primary elections held in May, and most districts in the state (70%) do so. A few, however, are exempt from this standard and hold school board elections during November even-year general elections[52] or odd-year municipal elections (N.C. Gen. Stat. § 115C-35(a)).[53] In any case, school performance information was linked to the subsequent race in which a school board election was on the ballot.[54]

Table A.4: Elections Used in School-Voter Match

School Year	AYP Data Released	Election Date
2003-2004	7/21/2004	11/2/2004
2004-2005	7/21/2005	11/8/2005
2004-2005	7/21/2005	5/2/2006
2005-2006	7/21/2006	11/7/2006
2006-2007	7/21/2007	11/6/2007
2006-2007	7/21/2008	5/6/2008
2007-2008	7/21/2008	11/4/2008
2008-2009	7/21/2009	11/3/2009
2008-2009	7/21/2009	5/4/2010
2009-2010	7/21/2010	11/2/2010
2010-2011	7/21/2011	11/8/2011
2010-2011	7/21/2011	5/8/2012

Note: AYP status is matched to the next election with a school-board race on the ballot. In North Carolina, school board elections are held at different times across school districts, but consistently within the same district over time. Thus, school performance for a given year in a given district is matched to only one election—the one of which there is a school board race on the ballot.

[52] The counties that hold their elections at this time include Alamance-Burlington, Alexander, Buncombe, Caldwell, Camden, Catawba, Cherokee, Cumberland, Davidson, Elkin City, Gaston, Guilford, Harnett, Haywood, Hertford, Hickory City, Hoke, Johnston, Lincoln, Macon, Nash-Rocky Mount, Rockingham, and Transylvania.

[53] The counties that hold their elections at this time include Burke, Chapel Hill-Carrboro, Charlotte-Mecklenburg, Cleveland, Mooresville, Newton-Conover, and Wake districts.

[54] Drilling down to the exact number of votes cast in a school board race is difficult given the limited data available in the state on local elections. This information is highly disaggregated and, as such, varies highly in its quality and content. Simply put, many local elections are missing in board counts and many others lack basic information such as candidate incumbency. For these reasons, the voter turnout measure used here constitutes the best measure available of voice-based responses in the state.

IV. Geographic Matching

As mentioned in the text, in order to match individuals in the voter file and schools, voters were matched to the school that the minimized Euclidean distance between the voter and a public school.[55] While this approach allows a scope that previous matches between voters and schools have rarely achieved, the disadvantage of this matching approach is that it introduces additional error into the data process.

Fortunately, this error is unlikely to introduce bias in the estimates. This is the case as, in North Carolina, geographic matching does reasonably well at assigning voters to assigned schools. Table A.5 shows this by comparing the geographic matching procedure to matching based on actual school assignment for a sample of voters. The sample of voters is determined based on the data available in the School and Boundary Information System (SABINS) housed at the University of Minnesota.[56] This database is only available for a handful of North Carolina school zones and only in one year (2010); hence, we cannot use it to match all, or even a significant portion of voters to schools. (Mismatch from the SABINS data is overstated as the available data comes primarily from the urban districts in North Carolina. Sparser, more rural districts, have fewer boundaries, thus school matching is likely to be higher across the whole state.) That said, this data can be used as an informative check of alignment between geographic and actual school-voter matching.

Overall, geographic matching puts voters in one of their assigned schools around 60% of the time, with voters in assigned districts 96% of the time. Additionally, mismatched and matches voters have similar voting status, race, party, gender, and age (the only individual-level factors available in the voter file).

Most importantly, this validation sample allows us to compare whether voters were matched to an assigned school at equal rates on either side of the school failure RDD. To verify this, we run the regression discontinuity models on an indicator variable for whether the geographic match aligned with an assignment match. If matching biased our results, we would expect to see a discontinuity in predicted match rates at the failure cutoff. We do not. The results reveal that elementary schools are balanced in match rates at the cutoff ($p=0.74$).[57] Hence, geographic matching is highly unlikely to bias the results presented. If anything, it simply makes it harder to see effects, if these are, indeed, present.

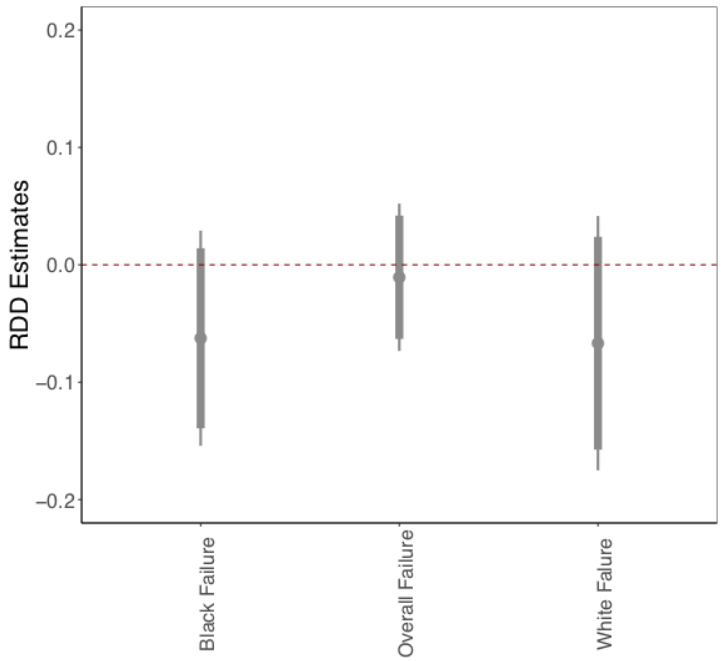
[55] Euclidean distances are measured "as a bird flies."
[56] The SABINS data is available at www.sabinsdata.org.
[57] This also holds for a verification match of the closest ($p=0.19$), middle ($p=0.38$), and high school ($p=0.32$).

Table A.5: Geographic vs. Assignment Matching

In Assigned School	58.3%	
In Assigned District	95.8%	
	In Assigned School	Not In Assigned School
%Active	78.0%	78.1%
%White	64.7%	69.7%
% Democrat	44.9%	42.0%
% Male	44.3%	44.3%
Average Age	47.11	48.07
p	0.74	

Notes: Geographic matching, though imperfect, does reasonably well at matching voters to their assigned school. This table compares geographic to assignment matching of voters to schools. Geographic matching pairs voters to their closest elementary school. Assignment matching pairs voters with the schools districts assign. Assignment matching is determined using the School and Boundary Information System (SABINS) housed at the University of Minnesota, which is only available for a handful of North Carolina districts only in 2010. The p-value reported near the bottom of the table is for the RDD test for imbalance in mismatch at the cutoff.

Figure A.4 Balanced Mismatch at the Cutoff



Notes: figure shows the balance in mismatch at the three cutoffs used in the paper. Observations are coded as being mismatched if they are not in the same school as they assigned to. Mismatch is determined by a match to the Sabins district boundary data for the school districts where this is available. None of the cutoffs show imbalances in mismatch (overall: p=0.74; white=0.23; black=0.182).

V. Additional Heterogeneities

Va. Sanction Status Heterogeneities

One might wonder whether the estimates we provided in the paper vary by the sanctions under NCLB. After all, being exposed to failure may vary depending on whether failure is a new phenomenon for a given community, or whether it has occurred repeatedly—thus putting the school under sanction. Table A.6 tests this possibility by interacting our treatment (school failure) by sanction status (which is determined by how many consecutive years a school has failed or not failed). As can be see, for overall failure signals there do not appear to be large heterogeneities in responsiveness by sanction status for the exit responses. White and black citizens respond the same regardless of how long a school overall has failed. However, for voting, failure messages amplify a response when a school has failed repeatedly. For each level that a school goes up the sanction chart, failure induces and additional 2.6 percentage point boost in school board turnout for white individuals ($p<0.05$) and 1.2 percentage point boost for black individuals ($p<0.10$). For race-based signals, it seems that the largest responses come earlier in the failure process. That is, individuals are more likely respond when the failure of their racial subgroup is a new phenomenon. All coefficients here are negative, with those for white exit, in-group; black exit, in-group; black exit, out-group; and black vote, out-group being statistically significant. All of these provide clear evidence that responses are lower in higher levels of failure. This deterioration of responsiveness may come for a variety of reasons. For example, they may occur because race signals attached to racial identities may be more salient (a fact consistent with our results in the paper that show larger effects for identity signals). This increased salience may mix with a depressive effect that comes from perpetual failure. When people recognize failure (i.e. when it is attached to a salient group), they may become used to school failure, thus making them less likely to respond. This finding is important as it runs counter to the hope of reformers: that perpetual failure would produce more citizen response.

Another point relevant to this discussion of heterogeneities along this dimension has to do with NCLB's specific transfer sanction. Under NCLB, once a school has failed twice consecutively, students are given the option to transfer to other schools in the district. It seems reasonable to wonder whether our exit options are driven entirely by the transfer option; that is, whether anyone else decides to exit by other means when in-district public school transfers are not in place—by moving, going to a magnet or charter school, going to private school, etc. (This possibility is most salient for exit; hence, we do not run this check for turnout.) In Figure A.5 below, we test this possibility. Here we show the exit estimates by sanction status—with schools being under sanction allowing transfers and those that are not under sanction not being able to do so.[58] The estimates show that exit is not entirely driven by the in-district public school transfer option. As can be seen, the failure estimates are similar across all conditions, with the exception of out-group failure for Black individuals (which shows an increase early in the process, but not when sanctions are in place—consistent with Table A.6). Across all other estimates, there appears to be an exit response even when transfers to other public schools in the district are not possible. That is, failure induces some families to find a way to move schools through other means other than the formal NCLB transfer application.

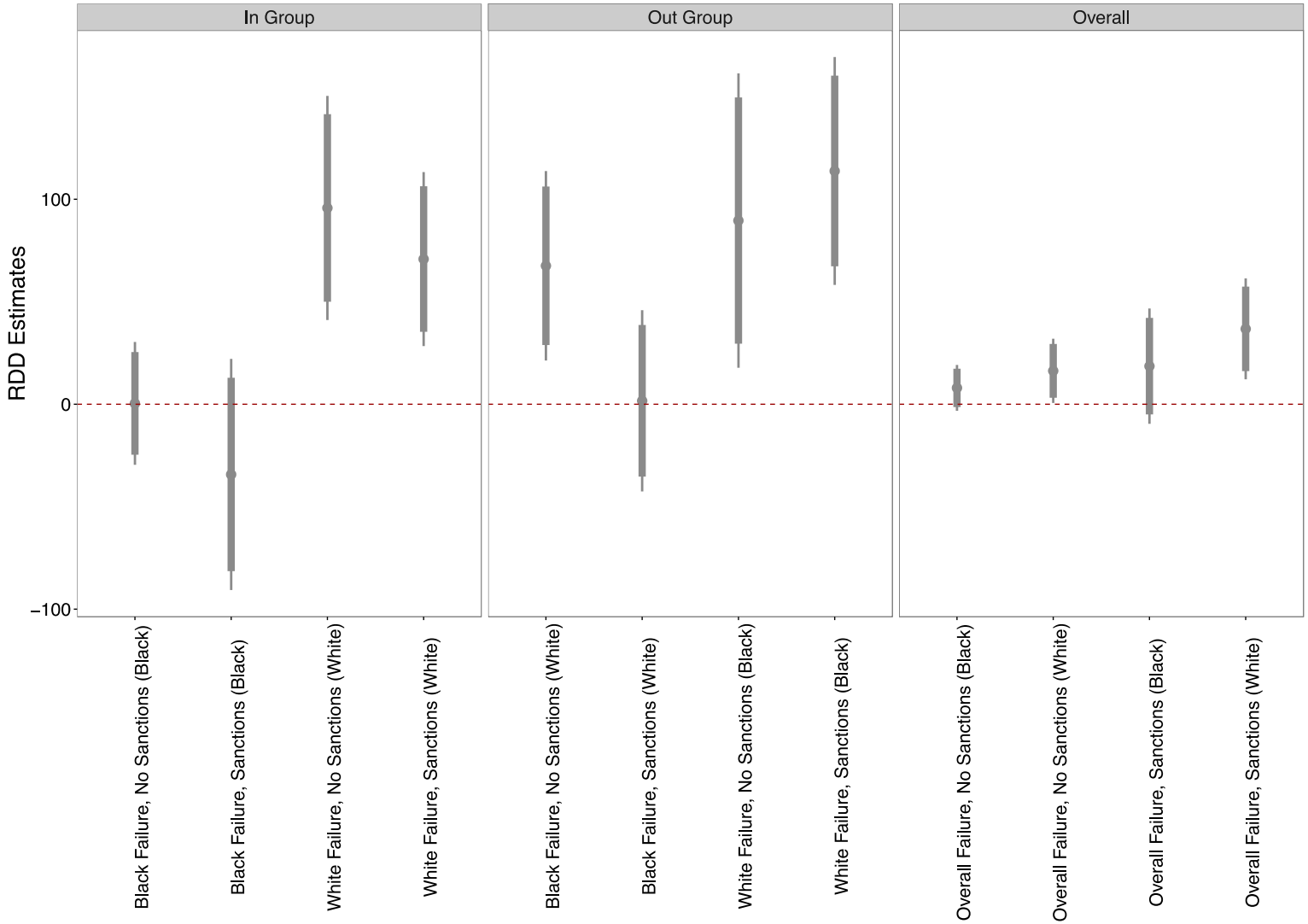
[58] It's important to note that this sub-setting (as opposed to interactions) of the data cuts our statistical power. As such, the key differences are across coefficients and in substantive terms.

Table A.6 Voice and Exit Estimates by Sanction Status

	Overall				In-Group				Out-Group			
	(1) DV: White Exit	(2) DV: Black Exit	(3) DV: White Vote	(4) DV: Black Vote	(5) DV: White Exit	(6) DV: Black Exit	(7) DV: White Vote	(8) DV: Black Vote	(9) DV: White Exit	(10) DV: Black Exit	(11) DV: White Vote	(12) DV: Black Vote
Failure	52.70***	41.13***	0.0269**	0.06***	117.1***	20.21	-0.0269	0.173***	90.75***	140.3***	0.160***	0.0269
	(6.089)	(6.623)	(0.0126)	(0.0136)	(13.76)	(16.50)	(0.0335)	(0.0260)	(20.026)	(17.579)	(0.0351)	(0.0523)
Sanction Level	16.15***	26.11***	-0.02***	-0.00559	27.92***	69.90***	0.00382	0.0223*	26.56***	37.77***	0.0143	0.0171**
	(2.527)	(4.765)	(0.00565)	(0.00470)	(2.154)	(9.534)	(0.00643)	(0.0134)	(4.959)	(3.452)	(0.0156)	(0.00833)
Failure * Sanction Level	3.438	5.975	0.026***	0.0119*	-27.4***	-32.5***	-0.0179	-0.0168	-1.858	-29.2***	-0.00902	-0.07***
	(3.072)	(5.290)	(0.00795)	(0.00675)	(4.070)	(10.01)	(0.0203)	(0.0147)	(5.620)	(5.898)	(0.0174)	(0.0216)

*** p<0.01, ** p<0.05, * p<0.1. Cluster-robust standard error-produced confidence intervals. Running Variable: flexibly-linear. Bandwidth: full. Each column represents a different regression discontinuity model. The number of students in each of the corresponding subgroups acts as a weight for these models. In the racial signal models, controls include the corresponding racial failure statuses that are not being tested.

Figure A.5 Exit Estimates by Sanction Status



Vb. Group Context Heterogeneities

For whites, we do see what seems to be a moderating role of context on the turnout of white voters as a result of white student failure, although the differences in effect estimates still do not quite reach statistical significance. As seen in Table A.7, in homogeneous white schools the failure of white students increases voter turnout considerably—nearly twice as much as it does among moderately white populations and nearly five times as large as areas with low white concentrations. While this difference is not statistically significant, it is substantively noticeable. A similar pattern exists with out-group signals—with most of the mobilization observed in Table 2 being a result of increased mobilization in high-white areas. The exact reason for this heterogeneity is unclear, but it may perhaps arise because dense white networks are able to better distribute failure signals among concerned parties, regardless of their origin. We think it important to note here that high propensity white schools are on the high end of the income distribution. Among white families, being in a homogenous, affluent social network makes it more likely that there will be a response in the voice domain.

Table A.7: Voter Turnout Response by Racial Context (White Citizens)

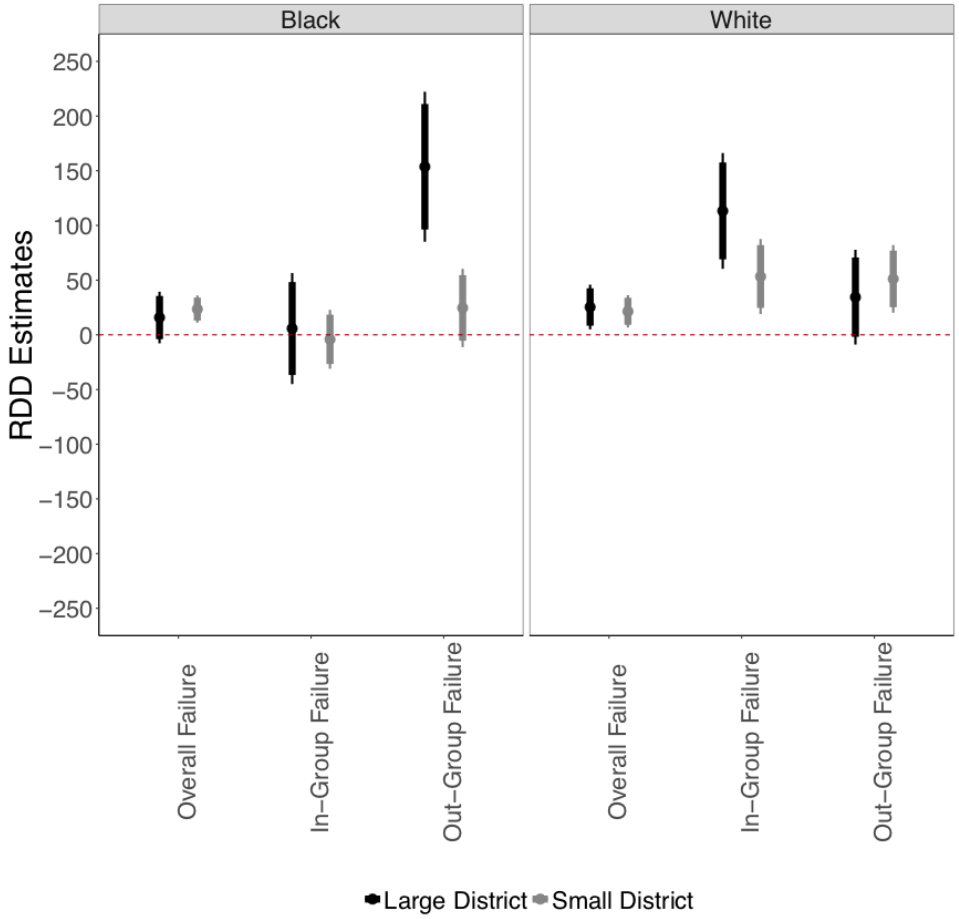
	(1) DV: White Vote (Low White)	(2) DV: White Vote (Medium White)	(3) DV: White Vote (High White)
Overall School Failure Signal	0.011 [-0.031, 0.054]	0.046 [-0.002, 0.094]	0.018 [-0.028, 0.063]
White Failure	0.032 [-0.085, 0.150]	0.071 [-0.030, 0.171]	0.148 [0.023, 0.273]
Black Failure	0.063 [-0.010, 0.136]	0.107 [0.001, 0.213]	0.251 [-0.026, 0.527]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. Racial groups are evenly divided across the three groups, with white in the minority representing schools with less than 42.0% in the school, racially mixed representing between 42.0% and 72.2%, and whites in the majority above 72.2%. In the racial signal models, controls include the corresponding White, Hispanic, or Black failure statuses that are not being tested. In–group signals are highlighted.

Vc. District Size Exit Heterogeneities

In Figure A.6 we display our exit estimates by district size; that is, by the number of schools in a district. Some might hypothesize that in districts with more schools, exit reactions may be larger, as the options available are larger for potential exiting families. In practice, we observe stability in our exit estimates across 4/6 of the treatments. The remaining two—out-group signals for Black individuals and in-group signals for white individuals—show signs of a larger exit response when there are more school options in a district. However, the difference is most stark for Black students. For Black students, with fewer exit opportunities and race as a whole, this may occur because transfers can only happen when many options are readily available. For white students, this difference is muted as exit options are more readily available to these more advantaged students.

Figure A.6 Exit Estimates by District Size



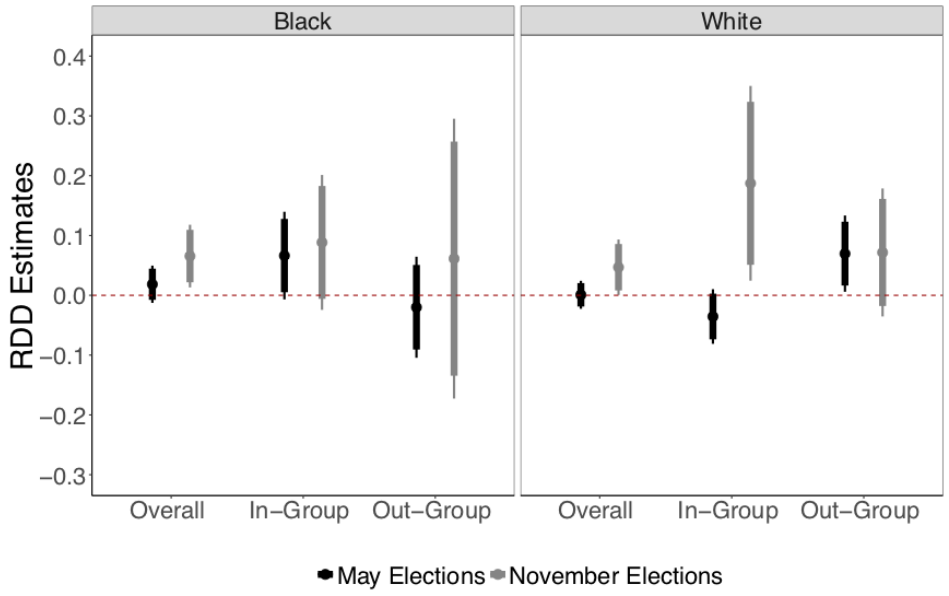
Notes: figure shows the exit effects broken by school district size. Coefficients are shown with a dot; 90% and 95% confidence intervals are shown with plots. Each coefficient plot correspond to a single RDD model. Running variable: nonparametric, Bandwidth: full. The two cells are broken by white and black individuals. The scale of the y-axis is in the number of children exiting a given school.

Vc. Election Timing Turnout Heterogeneities

As we mentioned in the text, school board elections occur at different times in North Carolina, with some occurring in May (70.7% of districts, 53.4% of schools) and the rest occurring in November. This places natural variation between how long between when the school failure signal is released (the end of July) and when the school board election occurs. One might wonder whether this matters for the responsiveness of the electorate. Ex-ante it is unclear whether it will. On the one hand, signals' effect may be magnified when performance information is released closer to the election. On the other, signals' effect may be the same given that these failure signals take time to diffuse and are held up by other institutions that keep them salient over time (home values, homebuyers, realtors, school board challengers, etc.; see Holbein 2016 for examples).

We test this in Figure A.7. As can be seen, across all treatment estimates, the effects are larger when the time between the release of the school performance signal and the election is smaller. However, in only 3 out of the 6 tests (overall failure, white; overall failure, black; in-group, white) is the difference statistically significant. Still, this suggests that timing may play a role in influencing responsiveness in the voice realm.

Figure A.7 Turnout Estimates by Election Timing



Notes: figure shows the turnout effects broken by the timing of school board elections. Coefficients are shown with a dot; 90% and 95% confidence intervals are shown with plots. Each coefficient plot correspond to a single RDD model. Running variable: nonparametric, Bandwidth: full. The two cells are broken by white and black individuals. The scale of the y-axis is in the RDD estimate in the proportion increase in school board turnout.

VI. Alternate Model Specifications

Table A.8 shows our exit estimates removing certain kinds of exit least likely to be linked to school failure—providing us some purchase on the types of exit that students use to respond to

failure. This provides us with a more precise measure of our exit outcome. Comparing Tables 2 and A.8 reveals that most failure-induced exits occur as in-district transfers. This can be seen by the fact that none of the significance levels or substantive sizes of the coefficients being meaningfully different.

Table A.9: NCLB Signals' Effect on Citizen Exit (Table 3 with no Out of District, Out of State, Private School Exits)

	(1) DV: White Exit	(2) DV: Black Exit
Overall School Failure Signal	21.78 [9.65, 33.91]	17.15 [2.77, 31.54]
White Failure	76.79 [47.19, 105.65]	91.27 [53.38, 129.16]
Black Failure	34.01 [4.85, 63.12]	-1.96 [-33.94, 20.14]

Cluster-robust standard error-produced confidence intervals. Running Variable: local-linear. Bandwidth: full. Each cell represents a different regression discontinuity model. The number of voters in each of the corresponding subgroups acts as a weight for these models. In the racial signal models, controls include the corresponding racial failure statuses that are not being tested. In–group signals are highlighted.

VII. Online Appendix References

Ahn, Thomas, and Jacob L. Vigdor. 2014. “When Incentives Matter Too Much: Explaining Significant Responses to Irrelevant Information.” *NBER Working Paper #20321*.

Chakrabarti, Rajashri. 2014. “Incentives and Responses under No Child Left Behind: Credible Threats and the Role of Competition.” *Journal of Public Economics* 110: 124-146.

Hemelt, Steven W. 2011. “Performance Effects of Failure to make Adequate Yearly Progress (AYP): Evidence from a Regression Discontinuity Framework.” *Economics of Education Review* 30: 702-723.

Holbein, John. "Left behind? Citizen responsiveness to government performance information." American Political Science Review 110, no. 2 (2016): 353-368.

Holbein, John B., and Helen F. Ladd. "Accountability pressure: Regression discontinuity estimates of how No Child Left Behind Influenced student behavior." Economics of Education Review 58 (2017): 55-67.

Traczynski, Jeff and Fruehwirth, Jane C. 2014. Spare the Rod? The Dynamic Effects of No Child Left Behind on Failing Schools *Working Paper.*