Online Appendix:
Left Behind? Citizen Responsiveness to Government Performance Information

This document contains supporting information for *Left Behind? Citizen Responsiveness to Government Performance Information* These are organized by commonly asked questions involving the paper text. Section I discusses how school failure signals are distributed. Section II explains why residents look to school boards instead of other civil servants or elected officials when schools fail. Section III documents the electoral contexts which school board elections occur. Section IV compares geographic matching to assignment matching using a limited sample where school zone assignment maps are available electronically. Section V gives an illustrative example of how the running variable in the regression discontinuity models used in the paper was determined. Section VI outlines validity tests for the failure cutoff. Section VII discusses whether exit biases the estimates presented in the paper. Section VIII provides information about the first stage estimates the in regression discontinuity models. Section VIX discusses whether the results are robust to alternate model specifications, providing visuals of the effects shown in the paper. Section X provides an alternate visualization of the quantile regression discontinuity results presented in the paper.

I. Failure Signal Distribution

Administrators are required to publicize AYP failure to the surrounding community. 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).[1]

In North Carolina, this requirement is fulfilled through two centralized channels: a state-run website and individual letters to those in the school-zone. Figure A.1 shows an example notification letter. The two centralized websites—www.accrpt.ncpublicschools.org and www.ncreportcards.org/src/—publish school failure status. Figure A.2 provides examples of the information that is available on the first of these two. The first panel shows the website interface; the second shows the format of information delivered for individual schools. Importantly, there is some evidence that these websites are widely used. As of July 2015, StatShow and SEMRush—two website traffic monitoring services—estimate that the first, broader site reaches roughly 100,000-140,000 unique visitors per month. The second site, which focuses solely on school report cards, but is more difficult to locate, generates visits from about 9,000-12,000 unique visitors per month. Additionally, there is some evidence that these sites receive the most traffic in the months after school performance is released. Still, the information on these websites is somewhat dense to navigate. This may contribute to why we see an effect among high propensity residents, but less so among low propensity. Consistent with this view, some have shown, in a lab setting, that making school performance data easier to interpret may increase the likelihood of a response (Jacobsen, Saultz, and Snyder 2012).

Beyond these two centralized channels, decentralized channels may work to distribute school failure to the broader community. There is some evidence that the media plays a key role in distributing school performance signals—for example, Berry and Howell (2007) provide some evidence that local newspapers frequently cover the release of AYP status. Additionally, the home-buying market may also play an important role: with highly informed realtors distributing AYP status to those purchasing homes

[1] Here, "corrective action" simply means failing to make AYP.

in the area (Black 1999; Figlio and Lucas 2004). These decentralized mechanisms likely combine with the centralized mechanisms to create the mobilizing effect described in the paper.

Figure A.1 Sample Adequate Yearly Progress Letter & Website Interface

[School District Logo] [School Name]

[School Address]

[Date]

Dear [School Name] Parent or Guardian:

I am writing to let you know that, according to [year] student assessment results, [school name] did not make Adequate Yearly Progress (AYP). [X]% of schools in our district made AYP for [year].

For the [next year] school year, if your school does not make AYP and your child is economically disadvantaged, he or she may be eligible to receive Supplemental Education Services (free tutoring).

We welcome your involvement at [school name]. There are many ways to participate including becoming a member of the Parent Teacher Association (PTA), offering your input to the School Improvement Team, volunteering, attending parent conferences with your child's teacher(s), making sure that your child is prepared and attends school each day, and monitoring your child's homework.

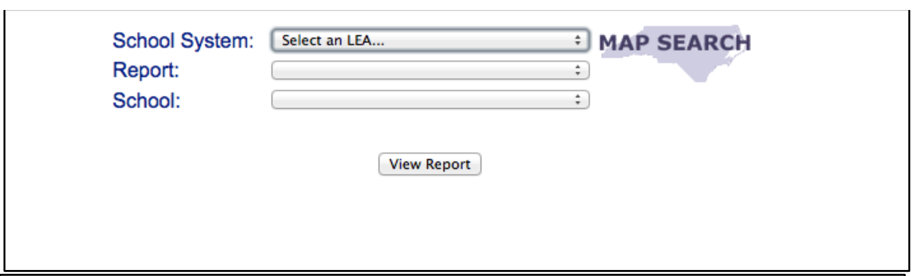
We are looking forward to a year of educational success and continuous improvement for all of our students. Please continue to work with us to make sure that we achieve success for each student.

Sincerely,

[Principal Signature]

[Superintendent Signature]

Information regarding NC School Report Cards can be found at [website].
Phone: (XXX)-XXX-XXXX Fax: (XXX)-XXX-XXXX



School Did Not Make Adequate Yearly Progress
School Met 18 (or 94.7%) out of 19 Target Goals

Reading Grades 3 through 8

	All Students	American Indian	Asian	Black	Hispanic	Two or More Races	White	Economically Disadvantaged	Limited English Proficient	Students With Disabilities
Number of Students	315	<5	<5	41	50	7	215	146	37	23
Number of Students Tested	315	<5	<5	41	50	7	215	146	37	23
Percent Tested	100%	*	*	100%	100%	*	100%	100%	*	*
Met 95% Target Goal?	Met	Insuf Data	Insuf Data	Met	Met	Insuf Data	Met	Met	Insuf Data	Insuf Data
Number of Tested Students (Full Academic Year)	305	<5	<5	39	45	7	212	137	35	20
Target Goal Percent Proficient (At or Above Grade Level)	71.6%	71.6%	71.6%	71.6%	71.6%	71.6%	71.6%	71.6%	71.6%	71.6%
Percent Proficient (At or Above Grade Level)	73.4%	*	*	*	51.1%	*	80.7%	59.9%	*	*
Percent Proficient with Growth	77.4%	*	*	*	60.0%	*	84.0%	67.9%	*	*
Met AYP Proficiency Goal?	Met	Insuf Data	Insuf Data	Insuf Data	Not Met	Insuf Data	Met	Met w/SH	Insuf Data	Insuf Data
Number of Students Included in Growth	196	<5	<5	22	25	<5	143	81	17	6
Percent Met Growth Expectation	59.7%	*	*	*	*	*	62.9%	54.3%	*	*
OAI Attendance%	95.9%	*	*	*	96.3%	*	95.9%	95.7%	*	*
OAI Attendance Met?	Met	Insuf Data	Insuf Data	Insuf Data	Met	Insuf Data	Met	Met	Insuf Data	Insuf Data

* = not calculated due to insufficient data (less than 40 students in the subgroup)

Mathematics Grades 3 through 8

	All Students	American Indian	Asian	Black	Hispanic	Two or More Races	White	Economically Disadvantaged	Limited English Proficient	Students With Disabilities
Number of Students	315	<5	<5	41	50	7	215	146	37	23
Number of Students Tested	315	<5	<5	41	50	7	215	146	37	23
Percent Tested	100%	*	*	100%	100%	*	100%	100%	*	*
Met 95% Target Goal?	Met	Insuf Data	Insuf Data	Met	Met	Insuf Data	Met	Met	Insuf Data	Insuf Data
Number of Tested Students (Full Academic Year)	305	<5	<5	39	45	7	212	137	35	20
Target Goal Percent Proficient (At or Above Grade Level)	88.6%	88.6%	88.6%	88.6%	88.6%	88.6%	88.6%	88.6%	88.6%	88.6%
Percent Proficient (At or Above Grade Level)	88.9%	*	*	*	82.2%	*	92.9%	85.4%	*	*
Percent Proficient with Growth	89.2%	*	*	*	82.2%	*	92.9%	86.1%	*	*
Met AYP Proficiency Goal?	Met	Insuf Data	Insuf Data	Insuf Data	Met w/CI	Insuf Data	Met	Met w/SH	Insuf Data	Insuf Data
Number of Students Included in Growth	197	<5	<5	22	26	<5	143	82	18	7
Percent Met Growth Expectation	77.7%	*	*	*	*	*	80.4%	70.7%	*	*
OAI Attendance%	95.9%	*	*	*	96.3%	*	95.9%	95.7%	*	*
OAI Attendance Met?	Met	Insuf Data	Insuf Data	Insuf Data	Met	Insuf Data	Met	Met	Insuf Data	Insuf Data

* = not calculated due to insufficient data (less than 40 students in the subgroup)

II. Why School Boards?

Among citizens' options for exercising voice when schools fail, residents could play a part in influencing and/or choosing the local school governing bodies, such as the school board. School boards play an integral role in education reform. Their decisions are of critical import, forming the basis of education processes in classrooms and schools (Ehrensal and First 2008). As such, school board elections play a fundamental roll in the education process. These elections tend to be close, further pushing individuals toward this venue for voicing their displeasure of school deterioration (for an example, see Green and Gerber 2008, 2–3). Simply put, residents concerned with failing schools have powerful reason to focus their efforts on the local school board.

Further, the policy feedback literature has shown that the traceability of a policy to an elected official—or the ease of which an individual can connect a policy outcome to a policymaker—matters in whether their is a response by the mass public (Pierson 1993). School boards are the most traceable elected officials when it comes to school outcomes, because they are the closest elected body to most parents.

There is some qualitative evidence that residents trace school failure to school boards. In a survey of parents, 78% indicated that school boards were extremely responsible when school performance deteriorated.[2] In the same survey, 88% of residents indicated that they believed that voting in school board elections would have an affect on school performance.[3] Finally, residents look to the school board for answers and solutions when their school fails. In another survey, citizens where asked to evaluate various solutions to school failure. Instead of blaming teachers and principals, residents looked to school boards most frequently for solutions. For example, providing more local comprehensive support from local officials outside the school received 54% of responses—almost 3 times as much as the next closest option.[4]

---

[2] Source: Accountability For All: What Voters Want From Education Candidates, Jan 2002.
[3] Source: Accountability For All: What Voters Want From Education Candidates, Jan 2002.
[4] Source: 2010 Phi Delta Kappa/Gallup Poll on What Americans Said About the Public Schools.

Taken together, these figures suggest that school board races are the right venue to look for a citizen response. For these reasons I restrict voting to elections in which a school board race was on the ballot.[5] This leaves me with a snapshot of the participation decisions made by the mass public in theoretically relevant elections.

[5] I use "school board election" and "election with a school board race on the ballot" interchangeably.

III. What Elections are used?

Table 1 documents how school performance matches to school board elections. Legally, NC school districts are required to hold school board elections in even-year primary elections (held in May). Most districts in the state (> 70%) do so. Districts that are exempt from this standard and hold school board elections during November even-year general elections 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 districts. Those that hold board elections during odd-year municipal elections include: Burke, Chapel Hill-Carrboro, Charlotte-Mecklenburg, Cleveland, Mooresville, Newton-Conover, and Wake districts (N.C. Gen. Stat. § 115C-35(a)).

Given this mix of electoral contexts under which school board elections occur, schools were matched to the next election in which a school board race would be on the ballot. The elections used are documented below.

Table A.1: 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. In North Carolina, school board elections are held at different times across school districts, but consistently within the same district. Thus, school performance for a given year is matched to only one election.

IV. Does Geographic Matching Bias the Results?

School and voter data are collected at different levels. The unit of observation in the voter file is the individual, and the unit of observation in the accountability data is the school. In order to fit the two data sources together, I matched registered citizens to the school that the minimized Euclidean distance between the registered voter and a public school.[6] The matching process identified the closest elementary, middle, and high schools, along with the most proximate school among the three. Individual citizens could be matched to any public school that reported NCLB school performance.[7] This matching process was relatively efficient, with 96.4% of citizens in the voter file being matched to the school closest to their address. There was no difference in match rates across passing and failing schools. The geographic matching procedure does relatively well at matching citizens to their assigned schools. And when it does not, no observables predict mismatch (see supplemental information).

The advantage to this matching approach is that it links school performance with validated voter behavior, something not done previously. Previous work has focused on matching school performance to survey responses, focusing on self-reports of school attendance and evaluation rather than actual behavior (Chingos, Henderson, and West 2012; Jacobsen, Saultz, and Snyder 2012). My strategy gives us a rich picture of the effect of school performance and citizen voice. The disadvantage of this matching approach is that some matches are more accurate than others. Though this approach comes with this disadvantage it is likely of minimal concern, outweighed by the benefit of matching schools to verified voting behavior.

In North Carolina, geographic matching does reasonably well at assigning registered citizens to assigned schools. Table 2 compares the geographic matching procedure to matching based on actual school assignment for a sample of registered voters. Assignment matching is determined using the School and Boundary Information System (SABINS).[8] This database is only available for a handful of North

[6] Euclidean distances are measured as one would measure with a ruler: "as the bird flies."
[7] Charter schools are included.
[8] Housed at the University of Minnesota, SABINS data can be downloaded at www.sabinsdata.org.

Carolina school zones and only in one year (2010); hence, I cannot use it to match all, or even a significant portion of citizens to schools.[9] Still, it can be used as an informative check of alignment between geographic and actual school-citizen matching for the subsample of schools for which both methods are available. The panels in Table 2 report alignment between assignment and geographic matching.

In each of the four panels, the first two lines report the percent of citizens placed in an assigned school and district by the geographic matching procedure. Overall, geographic matching puts citizens in one of their assigned schools around 60% of the time, with registered voters in assigned districts 90% of the time. Moreover, these estimates are likely biased downward as the available SABINS data comes primarily from the urban districts in North Carolina. Sparser, more rural districts have fewer boundaries, thus match rates are likely understating the effectiveness of the geographic matching.

The next five lines in Table A.2 compare registered voters that are matched to their assigned schools to registered voters that are not matched to an assigned school. Citizens where geographic matching aligns with assignment matching have similar voting status, race, party, gender, and age to those not in assigned schools. In short, when geographic and assignment matching are not consistent there are not observable voter traits that predict this distinction. As such, it appears that errors in matches are just adding noise to the estimates presented in the paper.

To confirm this assertion, I compare whether registered voters were matched to an assigned school at equal rates on either side of the school failure cutoff. To verify this, I reran the regression discontinuity models on a 0-1 variable for whether the geographic match aligned with an assignment match. If matching biased my results, we would expect to see a discontinuity in this variable at the failure cutoff. We do not. The results revealed that the closest (p=0.19), elementary (p=0.74), middle (p=0.38), and high school (p=0.32) geographic matches produced the same proportion of aligned matches on either

---

[9] Using SABINS data to match registered voters to schools for the regression discontinuity models is problematic because RD is a lower power method.

side of the failure cutoff. Thus, the geographic matching procedure I employ in the paper is highly unlikely to bias the results presented.

Table A.2: Geographic vs. Assignment Matching

	Closest	
In Assigned School	60.4%	
In Assigned District	92.4%	
	Assigned	Not
% Active	77.8%	78.2%
% White	63.8%	70.1%
% Democrat	45.4%	41.8%
% Male	44.2%	44.3%
Average Age	46.95	48.15

	Elementary	
In Assigned School	58.3%	
In Assigned District	95.8%	
	Assigned	Not
% 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

	Middle	
In Assigned School	52.3%	
In Assigned District	86.8%	
	Assigned	Not
% Active	77.5%	78.5%
% White	64.5%	70.5%
% Democrat	45.3%	41.3%
% Male	44.3%	44.2%
Average Age	46.89	48.39

	High	
In Assigned School	50.8%	
In Assigned District	87.2%	
	Assigned	Not
% Active	77.7%	78.4%
% White	62.2%	72.9%
% Democrat	46.5%	40.1%
% Male	44.0%	44.6%
Average Age	47.05	48.28

Notes: Geographic matching, though imperfect, does reasonably well at matching citizens to an assigned school. This table compares geographic to assignment matching of registered voters to schools. Geographic matching pairs citizens to their closest school (among all schools, elementary, middle, and high); while, assignment matching pairs citizens 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 and only in 2010. The panels report alignment between assignment and geographic matching when SABINS data is available. In each of the four panels, the first two lines report the % of geographically matched citizens in an assigned school and district. The next five lines compare registered voters in aligned schools vs. not. Alignment between the two matching methods is not predicted by observed registered voter traits.

V. How is the Running Variable Specified?

Table A.3 shows the inputs of treatment and the running variable. It shows an example of a single school. For this school, there are multiple subgroup scores repeated (up to 10 in each subject). For each subgroup, a determination is made based on whether the score is above or below the cutoff (marked by a horizontal line in each subgroup box).

As mentioned in the paper, this process follows the procedure of Ahn and Vigdor (2014a) and is meant to mirror the legal framework of NCLB. This approach is used because it bests others in its ability to correctly identify failure with the running variable, with ~80% of schools identified correctly. Misidentification comes from ambiguities in the proximity provisions.

Schools must be over the threshold in all subgroups to pass. School 1 fails to make AYP because of its performance among subgroups 7 and 10.

Table A.3: Channels and Subgroup Scores

School	Subgroup 1	Subgroup 5	Subgroup 7	**Subgroup 10**	Treatment & Proximity
School 1	**Proximity** Growth Level	**Growth** Proximity Level	**Level** Proximity Growth	**Level** Proximity Growth	Treatment: Fail Proximity: Subgroup 10

To identify the running variable, I choose one subgroup per school. In the example listed above, Subgroup 10's level score would identify the running variable. The intuition behind this choice is that in order to pass the school would have to raise Subgroup 7 and Subgroup 10 scores above the cutoff. Yet, as Subgroup 10 is further below the cutoff than Subgroup 7, this channel would have to be brought up the most in order for the school to pass. Thus, this gives a more accurate representation of how far a school is away from making AYP.

VI. Does the Discontinuity Satisfy Specification Checks?

If the assumptions of the RDD design hold, the estimate for Failing School (β_1) will be unbiased by confounders or simultaneity because schools fail as-good-as randomly within a narrow bandwidth (Lee 2008; Lemieux and Milligan 2008). Thus, the interpretations of coefficient estimates are similar to that in a randomized-control experiment. Determining whether the assumptions of RD hold, then, is of upmost importance. In this section I provide evidence for the validity of the failure to make AYP discontinuity.

The discontinuity at the failure cutoff could produce biased estimates if some other treatment shared the same cutoff as AYP. This is unlikely. To the author's knowledge, no previous research identifies NCLB's specific cutoffs as being substantively meaningful. The threshold for individual schools is sufficiently precise so as to allay any concern that the threshold itself had any substantive meaning. For example, elementary schools in 2010-11 had cutoffs of 71.6% (reading) and 88.6% (math) students proficient in the overall subgroup in order to pass AYP. These specific numbers have little meaning independent of their distinction as the pass/fail level. By all appearances they are arbitrarily set. Similarly, bias could be introduced if jumps were found at points near the cutoff. However, these are not found in my data. This is confirmed by placebo tests for jumps at points other than the discontinuity.

Additionally, if schools could *precisely* manipulate their proximity to failure score my results could be biased. Some have levied this criticism, for example, against the commonly used close-race discontinuities used to estimate electoral effects (Caughey and Sekhon 2011).[10] In the No Child Left Behind example, precise manipulation sufficient to threaten the validity of a discontinuity requires more than just a handful of schools being able to manipulate their scores. This type of behavior would have to be rampant, occurring differentially across relevant school characteristics (Lee and Lemieux 2009).

In order to check for precise sorting, methodologists argue that researchers "should begin by considering theoretical reasons for the violation of the RD assumption." In the NCLB example, precise

[10] Caughey and Sekhon argue that many variables are imbalanced in the close-election cut-point (2011). However, Eggers et al. show that this is an artifact of the modeling approach used and the data sources employed (2015).

sorting is theoretically unlikely for a couple of reasons. First, NCLB's usage of multiple subgroup categories makes it a difficult task for a large population of schools to manipulate their score. Doing so for one, let alone 20 subgroup categories is no easy feat. As such, precise sorting is not as straightforward in this regression discontinuity application. Second, if precise manipulation were to occur, a theoretical argument could be made that schools with higher resources would have the leg-up in the ability to do so. If this were the case, a disproportionate amount of high-resource schools at baseline would find themselves in the control group: able to ensure that they marginally pass AYP. Given the long empirical research showing that resources increase turnout across a variety of contexts, this would if anything provide conservative estimates of positive treatment effects. Put another way, if precise sorting occurred, effects that showed an increase in turnout in response to school failure would likely be biased downward by the artificially high number of high resource schools that make their way into the control group.

Additionally, if precise sorting occurred we would likely see it in empirical checks of covariate balance at the cutoff. In order to check for precise sorting—and other potential threats to a discontinuity—methodologists recommend two best practices. First, Eggers et al. recommend that researchers implement, "a battery of balance tests on pretreatment covariates and lagged values of the outcome variable, using the same specifications as the analysis on the outcome variable" (2015, 273). This test is similar to the covariate balance tests used in randomized control trials. If observable traits are relatively balanced at the cutoff, we can reasonably infer that the cutoff is arbitrary and that there is minimal precise manipulation of proximity to failure (Lee and Lemieux 2009). Second, methodologists also recommend the implementation of the McCrary Density test (2008)—a test that looks for clusters of schools centered on the preferred side of the treatment cutoff. The rationale here, applied to the NCLB example, is that if schools are precisely manipulating the failure cutoff, we might expect to see a cluster of schools that are just barely passing.

I implement and discuss both of these checks here. Both test support the validity of the school failure cutoff as sorting schools in an as-good-as random manner near the failure cutoff.

VIa. Covariate Balance

Table A.4 looks for imbalances among potentially influential covariates at the failure cutoff. This approach uses the same specification used to explore for treatment effects, reporting the coefficients and p-values from these models.

Of note in Table A.4 is that several potentially important confounders are balanced at the discontinuity. Important to the analysis at hand, all of the lagged dependent variables (in the first panel) are balanced at the cutoff. Substantively, this means that in the year before schools fail, they tend to have on average similar levels of voter turnout and exit. If precise sorting that biases the relationship between failure and our outcomes were to occur, we would likely see imbalances in these variables. We do not. In essence, this is a strongly informative placebo test for the validity of the discontinuity used (Eggers et al. 2015). That this discontinuity passes this test is very powerful evidence of the validity of the school failure discontinuity for the examination of these outcomes.

Table A.4: Balance of Baseline Covariates at Failure Margin

[1] Overview	[2] Variable	[3] T-C (β_1)	[4] P (T - C ≠ 0) | H_0
Lagged DV (Pre-treatment)	% Turnout	2.1%	0.31
	# Exit	2.37	0.85
	# White Transfers	-0.58	0.94
	# Black Transfers	1.26	0.77
	# Hispanic Transfers	0.42	0.74
	# Non-Econ. Disad. Transfers	-11.10	0.47
	# Econ. Disad. Transfers	0.90	0.89
School (Pre-treatment)	# of Students	-17.28	0.71
	# Transfer	-5.22	0.16
	# Transfer in LEA	-0.87	0.72
	# Transfer outside LEA	-1.72	0.18
	# Transfer out of state	-1.89	0.17
	# Transfer to private	0.04	0.93
	# Transfer to homeschool	-0.27	0.31
	% Female	-0.27	0.36
	Avg. Student Age	0.05	0.82
	% Migrant	0.02%	0.09
	% Parents with College	0.34%	0.74
	% Free/Red. Lunch	0.83%	0.27
	% Disabled	0.06%	0.30
	% African American	-0.22%	0.67
	# Missing Test	0.09	0.14
Voter Attributes (Pre-treat)	% Democrats	1.57%	0.17
	% African American	2.20%	0.02
	% Female	-.34%	0.65
	Age	-0.005	0.99
	# Registered voters	58.81	0.16
Post-Treatment Differences	Log Per-Pupil Expend	-0.30%	0.45
	Title I School	3.00%	0.13
	Principal Experience (Adv. Degree)	0.05%	0.92
	Principal Experience (11+ years)	0.46%	0.43
	Teacher Experience (% 11+ years)	-0.52%	0.06
	Teacher Experience (% Adv. Degree)	-0.57%	0.28
	Number of Teachers	-0.06	0.94

Note: Covariate balance at the failure discontinuity. The second column shows the variables explored. Column 3 shows the coefficient estimate for the regression discontinuity model used in the paper; column 4 shows the p-values from these coefficients. Data drawn from: NCERDC master-build, NCERDC end of grade, NCERDC end of course, NCERDC exit files, North Carolina accountability data, school district funding files, and the North Carolina voter file.

On other measures of potential importance at the school-year (second panel) and citizen-year (third panel) levels the story is the same.[11] Simply put, registered voters and schools near the cutoff are remarkably similar. Some variables are on the verge of being statistically distinct, but for the most part are substantively similar. Beyond the variables listed here, Holbein and Ladd (2015) explore whether there are imbalances in a larger set of school characteristics. They find that 23 out of the 26 variables (90%) examined are balanced at the failure cutoff.

Finally, it might be tempting to argue that other post-treatment differences may be driving the effect observed in the paper. In particular, if school failure led to changes in schools, then these may be behind the voice and exit effects shown in the paper. Simply put, people might be reacting to these downstream changes instead of the school failure signal, per-se. In this situation we would be measuring the indirect instead of the direct effect of the failure signal.

It is important to note that such a situation could explain why citizens respond to school failure signals—because failure sets in motion downstream consequences that residents don't like, and thus react to. As such, any indirect effects that failure elicits would still be meaningful: being a part of why NCLB elicits voice or exit responses. These would, however, change the interpretation of the effects list in the paper from being attributable to a failure signal to attributable to failure-induced changes. While it is impossible to account for all post-treatment differences—an inherent difficulty in standard approaches used to look for indirect effects (Imai et al. 2011; Green, Ha, and Bullock 2010)—I consider a few here.

If a school were to lose (or gain) money because it failed, citizens could be responding to these changes in funding rather than directly to the failure signal. This situation is unlikely given the provisions underlying NCLB and the timing of funding decisions. NCLB does not mandate any changes to funding rates if a school fails. These are generally set well in advance of school performance. By the time school failure status is determined in a given year—at the end of July—schools already have their funding levels

[11] In alternate model specifications checking for covariate balance, we see some evidence of imbalance in the proportion of females and the number of students in a school. To be abundantly cautious, these and the proportion African Americans are included as controls in the RDD models.

in place for the start of the next school year, which starts only a couple of weeks later. I show this in Table A.4. Though the estimate for per pupil expenditures in the next school year is negative, it is small (equivalent to $21 per pupil [p=0.51], or 0.02 of a standard deviation) and not statistically significant. In short, changes in funding rates are probably not driving the failure effects documented in the paper.

Similarly, it may be tempting to argue that the number of or attributes of school staff (i.e. teachers and principals) might influence voice and exit decisions. If failure moves these school attributes, residents may be react to these rather than the direct failure signal, per-se. However, it appears that these attributes are also locked in for the upcoming school year by the time failure status is determined. This is shown by the fact that teacher and principal post-treatment differences are all small and not statistically significant across failure status. In short, failure status is determined so close to the subsequent school year that teacher and principal characteristics are likely not driving the voice and exit effects observed.

In short, given the timing of the release of failure signals so close to the upcoming school year—as little as 2 weeks prior to the start of the school year—schools have little power to change many of their institutional features that residents might react to. As a result, residents are likely responding to the failure signal and not the downstream failure-induced changes. These post-treatment outcomes provide evidence that resource levels do not change before the elections held or documented exits occur. Hence, the effects documented in the paper are in response to school information signals and not changes in resources. Even after the elections occur and the exits in a given year are documented, most of the post treatment differences remain balanced. In the year after the effect is observed, Log Per-Pupil Expend (p=0.73), Principal Adv. Degree (p=0.34), Principal 11+ years Experience (p=0.88), Teacher 11+ years Experience (p=0.97), Teacher % Adv. Degree (p=0.92), Number of Teachers (p=0.08) all remain balanced. Title I School (p=0.02) shows some sign of imbalance, but this could be due to random chance associated with multiple hypothesis test run (we would expect about 0.5 of these 7 tests to be significance just by chance alone). Moreover, that this single imbalance occurs after the treatment effects are observed. This result strengthens the point that these are due to information, and not resource changes.

Based on this evidence it seems that school failure sorts schools into marginally passing and marginally failing categories in an as-good-as random manner. This discontinuity shows little evidence of shared treatment or precise manipulation of student performance. The specification tests outlined above support the use of the school failure cutoff as sorting schools in an as good as random fashion. Thus, we can likely use failure status as a test of citizens' responsiveness to the performance based accountability signals.

VIb. McCrary Density

To support the validity of the failure discontinuity, I also provide results from the McCrary density test for precise sorting at the cutoff. In this test if there is jump in the running variable at the discontinuity point, it is possible that unobserved manipulation of the running variable is occurring (Lee and Lemieux 2009). The McCrary density test provides another empirical check of local randomization, though it has been argued to be inferior in some regards to covariate balance as a test of precise manipulation (Lee and Lemieux 2009).

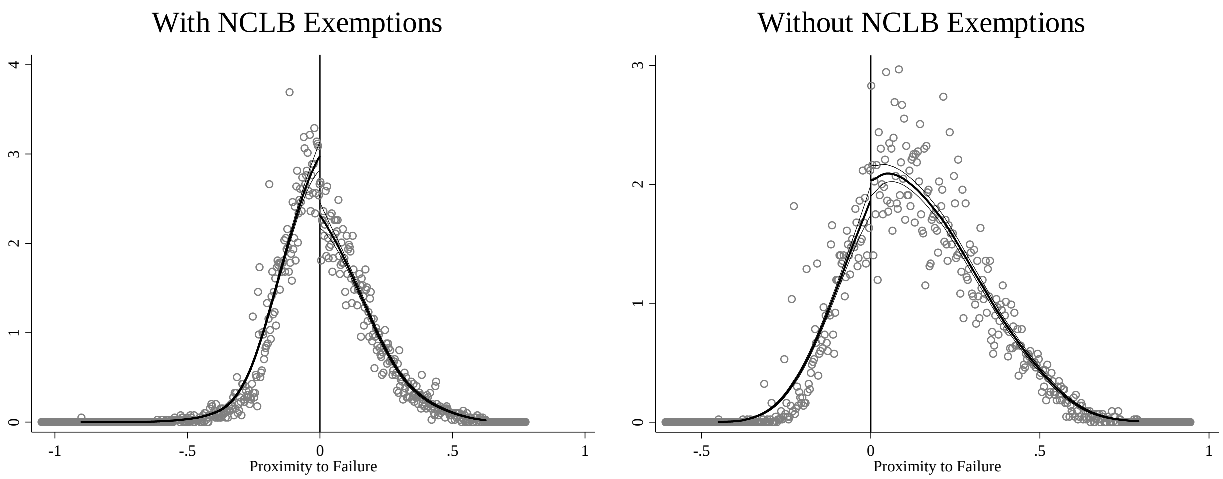
Figure A.3 shows McCrary's test for a discontinuity in the running variable. For the specification of the running variable used in the paper, there is a cluster of schools just on the marginally passing side of the AYP cutoff. This means that the school failure discontinuity marginally fails the McCrary check (p=0.04).

However, three pieces of evidence support that this may simply reflect the underlying distribution of school performance, rather than precise manipulation of the running variable. First, the cluster of schools disappears when the running variable is specified as the average subgroup score. Second, if I break down the density test by year (not shown), most of the years satisfy McCrary's suggestive conditions, with no clear pattern of schools increasingly being able to game the system over time. There is no evidence of schools learning how to manipulate the failure cutoff over time. More often than not, schools are sort as-good-as randomly at the discontinuity threshold (Ahn and Vigdor 2014a; Holbein and Ladd 2015). Finally, and most importantly, the balance of observable characteristics at the margin makes it unlikely that rampant precise manipulation of proximity to failure occurs. If schools are able to rampantly manipulate the failure cutoff, we are not able to see it, as we might expect, in their observable characteristics. That we don't is a powerful support that the discontinuity is valid.

But what explains this break at the cutoff? In all likelihood, it has to do with the exemptions granted to individual schools under NCLB. As discussed in the text, some schools are brought above the failure cutoff based on growth or "confidence interval" exemptions. Others have noted that these exemptions are likely behind the distribution of schools around the failure cutoff, rather than schools' ability to precisely manipulate the running variable (Traczynski and Fruehwirth 2014; Holbein and Ladd 2015). They show this by documenting that the cluster of schools just above the failure cutoff disappears when the confidence interval and growth exemptions are left out of the running variable calculation. I replicate this finding in Figure A.3 below by recalculating the running variable without these exemptions. True enough, the cluster of just passing schools disappears under this specification of the running variable. This suggests that the distribution of schools around the cutoff is driven by policy-driven exemptions rather than the precise manipulation of schools. This, reinforced with the other theoretical reasons and empirical checks, offers powerful evidence that the failure cutoff is sorting schools in an as-good-as random manner.

Still, to be abundantly cautious against the possibility of some other underlying trait driving precise manipulation, I include the school fixed effect mentioned in the text in the regression discontinuity models. This should absorb schools "propensity to manipulate" the treatment as long as this ability remains constant over time. This combined with covariate balance at the cutoff offer reassurance to the marginal failure of this informal specification test.

Figure A.3: McCrary Density Check



Notes: McCrary density test (2008) for precise sorting at the failure cutoff. The left panel shows the check with the distribution of the running variable incorporating NCLB's exemptions (confidence interval and growth). It shows a group of schools just marginally making AYP. The right panel shows the distribution of the running variable when the exemptions are not incorporated. The group of schools just marginally making AYP disappears. This second panel reveals that the cluster on the just passing side of the cutoff is due to the exemptions NCLB grants, not due to precise sorting of schools across the failure cutoff. In this case, failure of the McCrary test appears to be a product of the specification of the running variable, not a process of learning to game the system.

VII. Are the Turnout Effects Confounded by Exit?

The estimates in the paper indicate that failing schools increase voter turnout in elections involving a school board race. However, when individuals receive a signal that their school has failed, some choose to vote with their feet rather than voting at the ballot box. This behavior could drive or confound my results. Indeed, a decline in the denominator might be behind the rise in overall turnout.

Two factors determine the nature of this bias: first, the extent to which exit occurs and second, the attributes of those who leave. If we can identify the type of individuals who exit when a school fails, we can sign the bias introduced to the relationship between school failure and voter turnout. I examine these two factors here.

First, the models from the paper indicate that when schools fail, citizens react by voting with their feet. However net movements are not substantively large, unlikely to produce increases in turnout alone (by decreasing the denominator of voter turnout). A decrease on the order estimated in the paper—though it has implications for local schools—is not large enough to produce the substantive increases documented in the paper. Second, though net movements are small, movement of certain types of citizens may play a large role, introducing bias into the failure-turnout relationship estimated in the paper. Table A.5 demonstrates the nature of the bias depending on the type of individuals who move. If high propensity voters (high income, education, etc.) move to passing school zones when their school fails, my estimates would be biased downward. More high propensity voters in the control group (passing schools) would inflate their turnout numbers in subsequent elections, thus narrowing the effect size of failure on turnout. This type of outcome is documented in cell [2].

Table A.5: Bias in Failure-Turnout Estimates from Movers

Voter Type	Move from a Passing to a Failing School	Move from a Failing to a Passing School
High Propensity	[1] Overestimate	[2] Underestimate
Low Propensity	[3] Underestimate	[4] Overestimate

Conversely, if low propensity voters move from failing schools to passing schools—as shown in cell

[4]—I would overestimate my effects. This second type of behavior would be particularly troubling.[12] However, models in the paper indicate that high propensity voters: those affluent and Caucasian are more likely to exit than poor, minority individuals. Thus, failure causes high propensity voters to move from failing schools to passing schools, putting us in cell [2]. If anything, exit likely causes a slight underestimate of failure's effect on turnout. Put differently, by encouraging exit, failure signals undercut voice in school zones by giving an alternate venue for high propensity voters to express their discontent.

In sum, there is reason to believe that my estimates are close to the true effect of failure on turnout, if anything, biased slightly downward. When schools fail, movements occur, but not large enough to drive the voice results. Moreover, when schools fail those who have more education and income exit.

[12] The other two scenarios, 1.) High propensity voters moving to failing schools and 2.) Low propensity voters moving to passing schools, are unlikely or not likely to nullify my positive result in the paper. Scenario 1 would require a highly counterintuitive result, in which highly educated voters move from passing schools to failing schools. Scenario 2 may occur, but likely biases failure-turnout relationship downward.

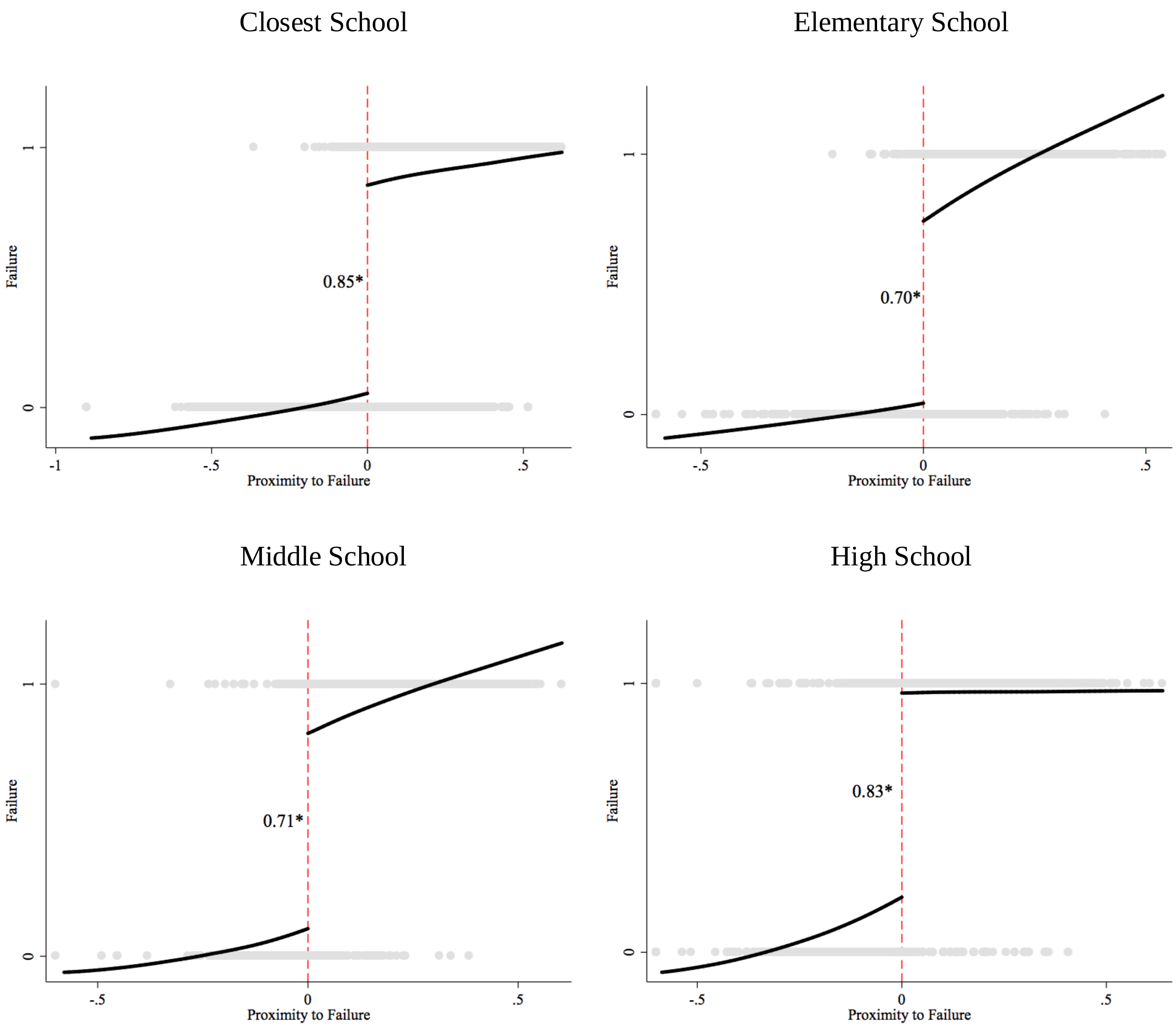
VIII. What About the First Stage Estimates?

As the models employ fuzzy regression discontinuity, it is reasonable to wonder about the first stage estimates. Simply put, is the instrument produced by the running variable strong enough? Or, put differently, is the running variable producing a sufficiently large jump in the probability of failure at the failure cutoff? To address this, I produce graphs showing the first stage estimates, shown in figure A.4 below. Each of these show that the running variable used produces a strong instrument.

Figure A.4 is based on non-parametric specifications of the running variable and the full bandwidth. These subfigures show the first stage estimate—that is, the increase in the probability of failure at the failure cutoff. Each of the panels shows that the Ahn & Vigdor (2014a) running variable identifies schools as passing/failing very well. The jump in the probability of failure at the cutoff is 0.85 for the closest school match, regardless of school level; 0.70 for the closest elementary school; 0.71 for the closest middle school; and 0.83 for the closest high school. All of these estimates are highly significant, easily clearing the 1% significance threshold. These estimates are also substantively large, with each of these easily clearing the recommended thresholds for instrument strength provided by Stock and Yogo (2005). These conclusions all hold with parametric specifications of the running variable.[13] They also hold for the district-level estimates—with this collapsed discontinuity producing a discrete jump in the number of failing schools that is sufficiently large and statistically meaningful.

[13] Non-parametric first stage estimates (with corresponding standard errors) are: 0.78* (0.00) for the closest school, regardless of level; 0.63* (0.00) for elementary schools; 0.68* (0.00) for middle schools; and 0.78* (0.00) for high schools.

Figure A.4 First Stage Estimates



VIX. Are the Results Robust to Alternative RD Specifications?

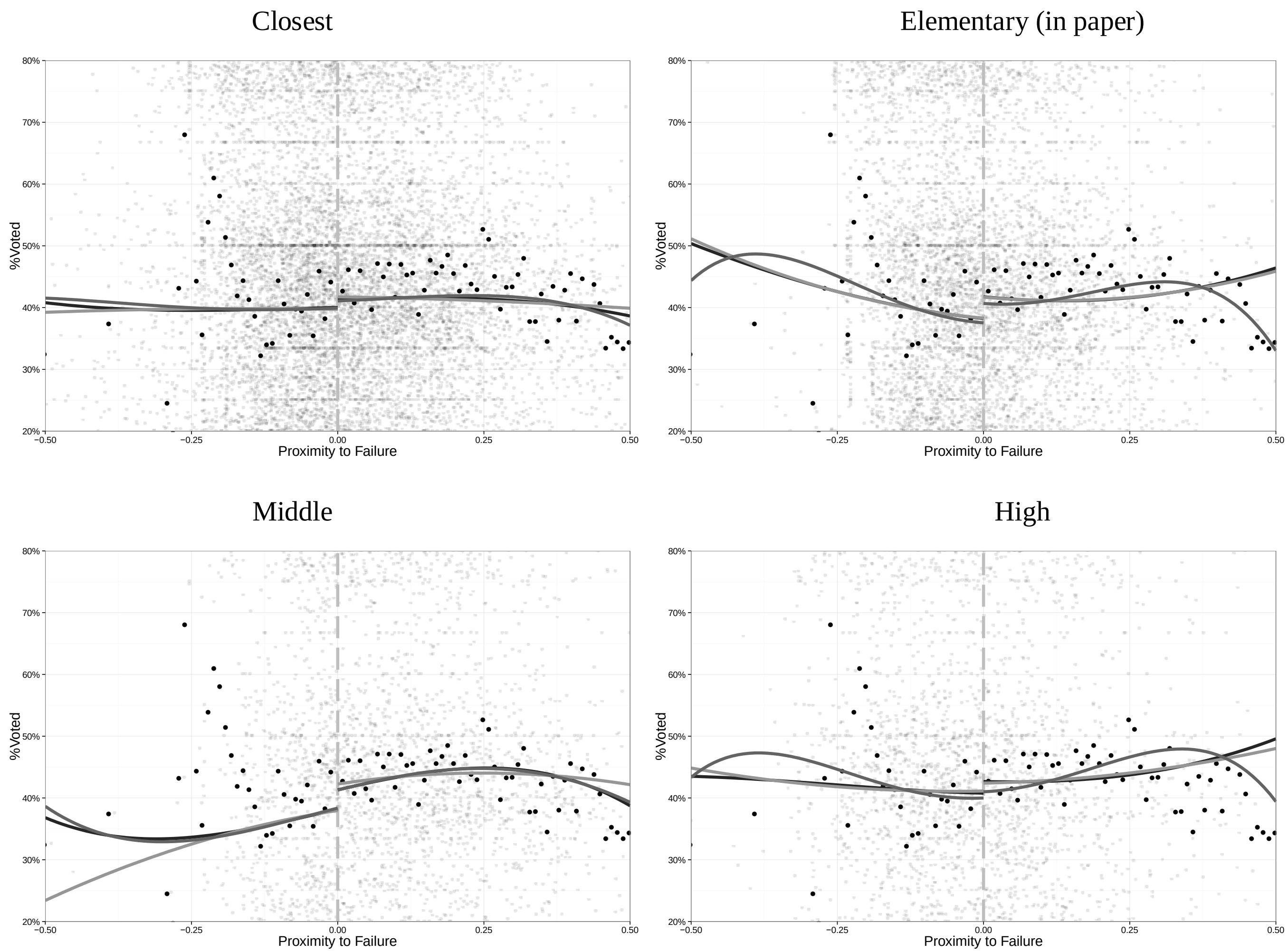
The results shown in the paper are robust to the bandwidth selection (to a point, given the finite number of schools narrow bandwidths are low power given a finite number of schools) and alternative specifications of the running variable. Table A.6 shows specifications based on various model specifications. These show consistent positive turnout estimates, with a few instances where statistical precision is not possible at traditional levels. However, substantively, the estimates remain in the same ball-park across estimates. Figure A.5 plots a few of these variations, focusing on alternate specifications of the running variable.

Table A.6: Alternate Specifications School Failure & Turnout

	[1] DV: Prop. Voted (Closest)	[2] DV: Prop. Voted (Elementary)	[3] DV: Prop. Voted (Middle)	[4] DV: Prop. Voted (High)
Linear, FE, Controls	.054*	.046*	.098*	.042*
	[.036, .71]	[.021, .072]	[.059, .137]	[.001, .083]
Quadratic, FE, Controls	.046*	.054*	.083*	.042*
	[.029, .064]	[.029, .079]	[.036, .130]	[.001, .083]
Cubic, FE, Controls	.027*	.058*	.079*	.043*
	[.006, .049]	[.027, .089]	[.027, .132]	[.002, .084]
Quartic, FE, Controls (in paper)	.025*	.052*	.079*	.051*
	[.003, .047]	[.020, .083]	[.026, .132]	[.001, .102]
Quartic, FE, No Controls	.023*	.052*	.080*	.048*
	[.002, .045]	[.020, .083]	[.027, .133]	[.000, .096]
Quartic, No FE, No Controls	.005	.037	.048	.024
	[-.024, .035]	[-.006, .080]	[-.020, .117]	[-.041, .090]
Quartic, No FE, No controls, No Weights	.021*	.071*	.023	.018
	[.000, .044]	[.037, .105]	[-.028, .073]	[-.020, .056]
Linear Interaction, FE, Controls	.054*	.057*	.097*	.044*
	[.037, .072]	[.033, .080]	[.059, .136]	[.002, .087]
Quadratic Interaction, FE, Controls	.013	.047*	.080*	.043*
	[-.009, .035]	[.012, .081]	[.031, .129]	[.000, .086]
Non-Parametric, No FE, No controls, No Weights	.027*	.069*	.037	.011
	[.008, .046]	[.039, .098]	[-.009, .083]	[-.031, .053]

Notes: $^{+}$ < .10 * p < .05. 95% Confidence intervals in braces. Constant suppressed so as to include all fixed effects in models with fixed effects. *Unit of analysis*: school-year (voter weighted). Standard errors are clustered to school level. *Bandwidth*: full. *Running variable*: quartic. Models with *controls* include the % African American (voter file), % Female (voter file), and number of registered voters in school zone (voter file). Non-parametric models use the `rdrobust` command in Stata (Callonico, Cattaneo, and Titiunik 2014).

Figure A.5: Failure's Effect on Turnout Visualized

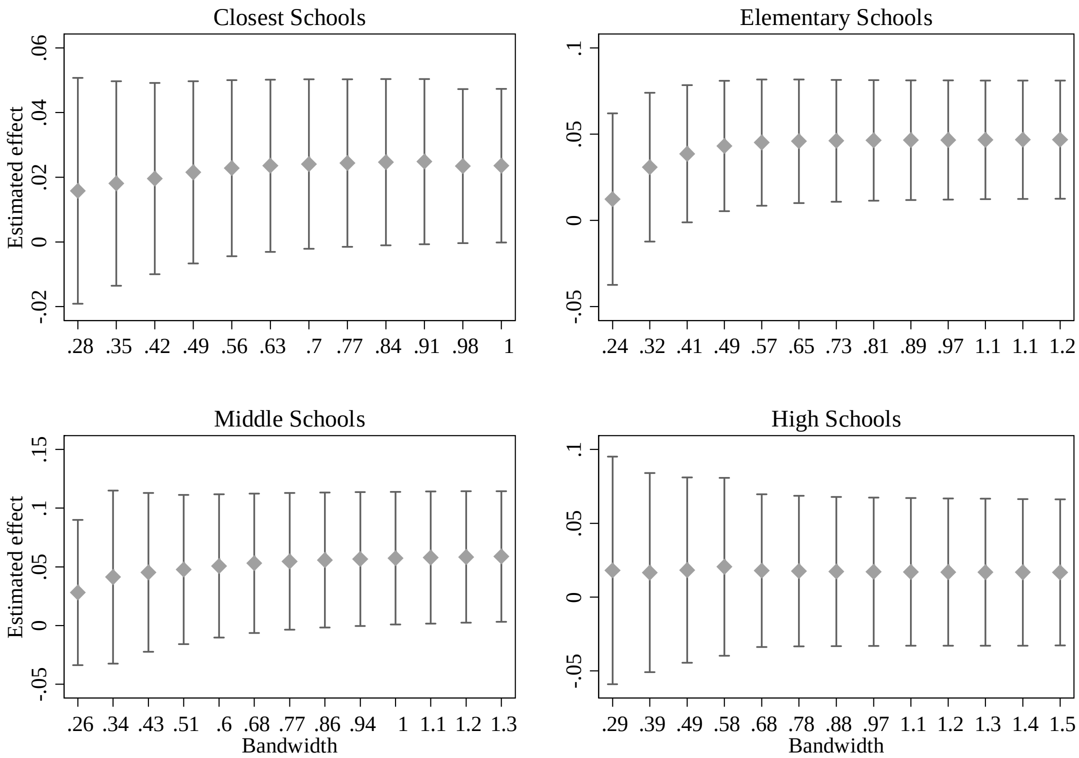


Points in the background display levels of turnout in the school zone, with binned averages **bolded**. Figure displays the causal effect of elementary school failure on the number exits from the school. Points in the background display the number of exits from a school, with binned averages **bolded**. The effect of failure is the distance between the corresponding lines. The figure demonstrates the effect is robust to alternate parameterizations of the running variable

Figure A.6 shows the regression discontinuity model estimates without fixed effects and a local linear regression of the running variable estimated separately on either side of the discontinuity. The results with a fixed effect and a linear interaction modeling of the running variable become more precise, and tend to show effects that are more consistently significant across bandwidths. The figures are oriented using the Imbens & Kalyanaraman (2012) optimal bandwidth, starting on the left with bandwidths of about 10 percentage points on either side of the discontinuity, to ensure a reasonable number of data points, and going all the way to full bandwidth of data on the right (60 percentage points on either side).

As we would expect, the estimates in the wider bandwidths are more precise. These benefit from having enough weighted school-year observations to precisely estimate the coefficients. Moreover, the estimates do not vary by bandwidth—wider bandwidths are statistically and substantively to those in the narrower bandwidths. Of all the four school estimates, the substantive size of the elementary school estimate varies the most across bandwidths. Still, these estimates are not statistically distinct across the bandwidths. Also, in some of the narrower bandwidths we do not have enough power precisely estimate the effect of failure signals on turnout. Still, it is assuring that the estimates are in the same neighborhood across the bandwidths, even when weighted school-year observations are less abundant. The lack of data may especially be behind the lack of significance when we look at the effect of high school failure. Compared to the other two school types, high schools are fewer in number and as such their estimates have less precision. As shown in the paper, high school estimates benefit from the inclusion of a school fixed effect.

Figure A.6: Failure's Effect on Turnout by Bandwidth



Bandwidth	Elem	Mid	High
0.2	4107	1635	1488
0.4 (≈ IK)	6718	2693	2524
0.6	7440	3224	2943
0.8	7604	3498	3072
1	7651	3603	3115
1.2 (full)	7661	3620	3123

Notes: Check across different bandwidths using non-parametric regression. Unit of analysis is the school-year, weighted by the number of students in the school. The graph orients itself with the Imbens & Kalyanaraman (2012) optimal bandwidth, which in the four models is approximately 0.4. The corresponding table on the right shows the number of school-year observations in the estimation sample.

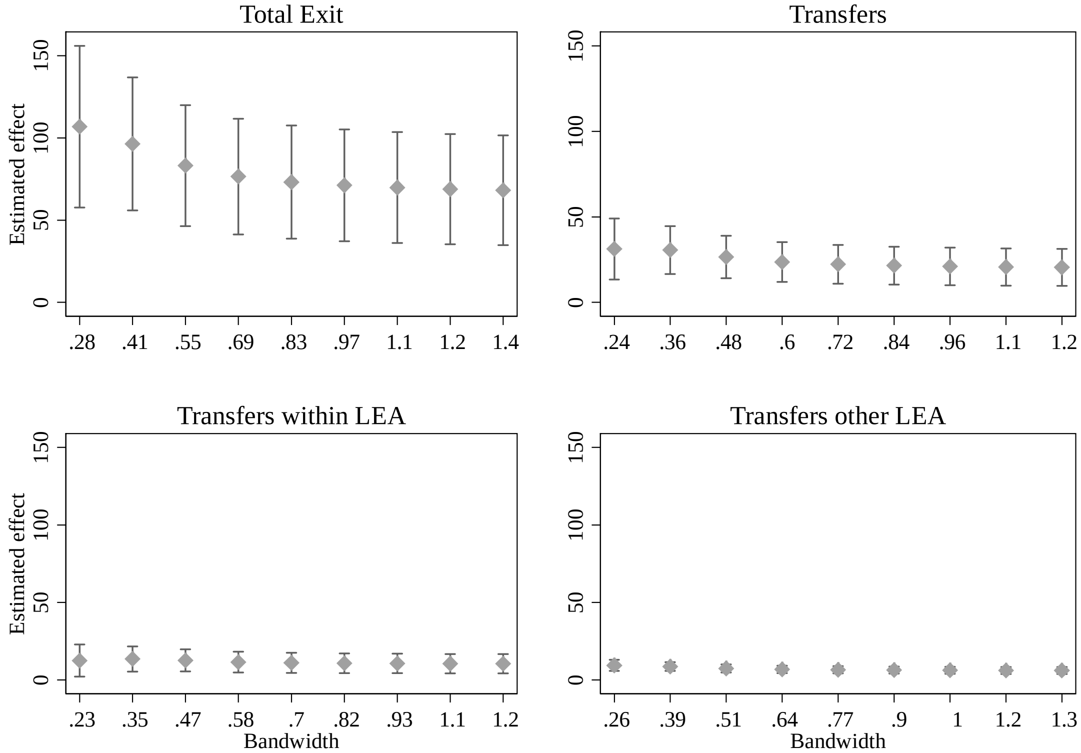
Table A.7 shows alternate parameterizations that show failure signals' affect on exit. Like the turnout estimates, these are remarkably similar across various model specifications. Figure A.8 shows the robustness of this result across bandwidths. In general, we are able to be much more precise with the exit estimates, with these are showing that failure increases exit.

Table A.7: Alternate Specifications School Failure & Exit

	[1] DV: # Exit	[3] DV: # Transfer	[4] DV: # Transfer in LEA	[5] DV: # Transfer Other LEA	[6] DV: # Transfer Out of State	[7] DV: # Transfer Private	[8] DV: # Transfer Homeschool
Linear, FE, Controls	38.13*	13.12*	8.32*	3.41*	-0.99	0.85*	0.44*
	[26.73, 49.55]	[7.42, 18.81]	[4.15, 12.49]	[2.04, 4.78]	[-2.40, 0.42]	[0.34, 1.36]	[0.22, 0.65]
Quadratic, FE, Controls	21.23*	14.45*	8.99*	3.83*	-0.99	0.98*	0.47*
	[11.14, 31.33]	[8.58, 20.31]	[4.73, 13.24]	[2.37, 5.29]	[-2.41, 0.43]	[0.46, 1.50]	[0.25, 0.69]
Cubic, FE, Controls	15.94*	17.66*	10.50*	4.88*	-1.01	1.32*	0.57*
	[0.80, 31.08]	[11.84, 23.49]	[6.18, 14.82]	[3.42, 6.33]	[-2.47, 0.46]	[0.83, 1.81]	[0.34, 0.80]
Quartic, FE, Controls (in paper)	15.99*	9.75*	4.94	2.96*	0.05	0.85*	0.35*
	[1.22, 30.78]	[1.66, 17.85]	[-0.66, 10.53]	[0.72, 5.21]	[-1.74, 1.85]	[0.29, 1.40]	[0.01, 0.69]
Quartic, FE, No Controls	12.06	18.69*	10.45*	5.20*	-0.74	1.51*	0.78*
	[-1.95, 26.08]	[12.59, 24.80]	[6.09, 14.81]	[3.67, 6.73]	[-2.20, 0.72]	[1.02, 1.99]	[0.51, 1.04]
Quartic, No FE, No Controls	105.01*	14.14*	8.82*	4.64*	-2.28	1.17*	0.53*
	[61.49, 148.53]	[4.08, 24.20]	[3.12, 14.52]	[2.52, 6.76]	[-5.15, 0.58]	[0.56, 1.78]	[0.26, 0.80]
Quartic, No FE, No controls, No	42.76*	5.54*	3.50*	1.59*	-0.60	0.42*	0.20*
Weights	[28.71, 56.82]	[2.12, 8.96]	[1.57, 5.42]	[0.76, 2.42]	[-1.49, 0.30]	[0.25, 0.59]	[0.10, 0.30]
Quadratic Interaction, FE, Controls	18.38*	10.52*	6.61*	2.45*	-0.29	0.67*	0.30
	[3.05, 33.72]	[2.48, 18.59]	[0.93, 12.30]	[.54, 4.36]	[-1.89, 1.31]	[0.03, 1.31]	[-0.01, 0.60]
Linear Interaction, FE, Controls	38.16*	12.90*	7.81*	3.44*	-0.83	0.93*	0.52*
	[27.68, 48.65]	[7.82, 17.99]	[4.15, 11.46]	[2.21, 4.67]	[-2.09, 0.42]	[0.48, 1.38]	[0.32, 0.73]
Non-Parametric, No FE, No	34.48*	8.46*	3.76*	2.36*	1.50*	0.32*	0.13*
controls, No Weights	[22.31, 46.65]	[4.53, 12.40]	[1.56, 5.97]	[1.41, 3.32]	[0.56, 2.44]	[0.11, 0.54]	[0.01, 0.25]

Notes: * p < .05. 95% Confidence intervals in braces. Constant suppressed so as to include all fixed effects in models with fixed effects. *Unit of analysis*: school-year (voter weighted). Standard errors are clustered to school level. *Bandwidth*: full range. *Running variable*: quartic. *Controls*: % African American (voter file), % Female (voter file), school size (school file). Non-parametric models use the `rdrobust` command in Stata (Callonico, Cattaneo, and Titiunik 2014).

Figure A.7: Failure's Effect on Exit by Bandwidth



Bandwidth	Elem	Mid	High
0.2	4107	1635	1488
0.4 (≈ IK)	6718	2693	2524
0.6	7440	3224	2943
0.8	7604	3498	3072
1	7651	3603	3115
1.2 (full)	7661	3620	3123

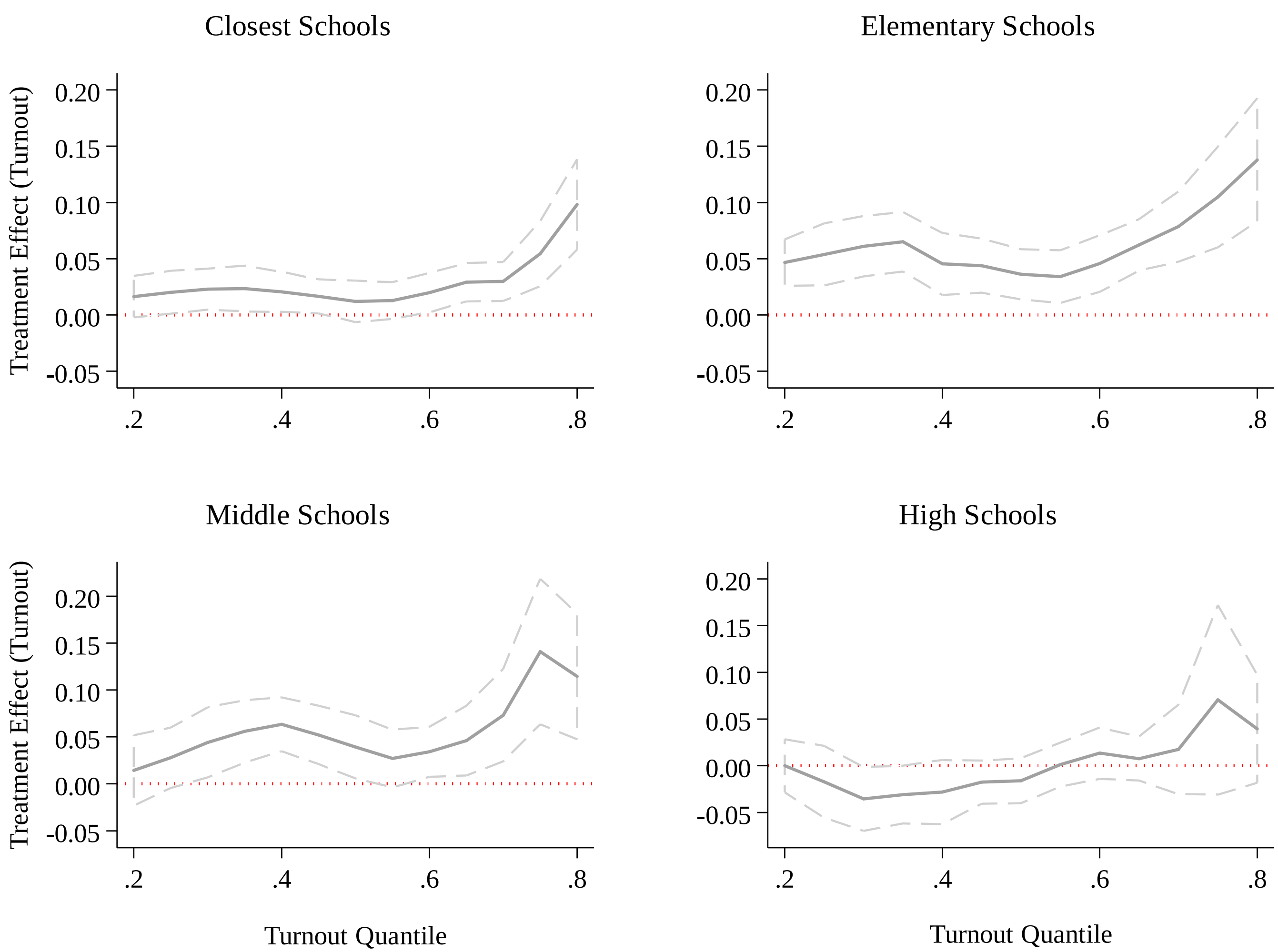
Notes: Check across different bandwidths using non-parametric regression. Unit of analysis is the school-year, weighted by the number of students in the school. The graph orients itself with the Imbens & Kalyanaraman (2012) optimal bandwidth, which in the four models is approximately 0.4. The corresponding table on the right shows the number of school-year observations in the estimation sample.

X. Quantile Regression Results

The results from the quantile regressions shown in the paper (Figures 2 and 3) are simplifications of the quantile models run, included for ease in interpreting this method that may be unfamiliar to many in political science. The figures in the paper do not account for uncertainty. As such, figure A.8 shows the results from these quantile regression models with the 95% confidence intervals for the estimates.

If school failure mobilized all groups equivalently the coefficients would be relatively similar across deciles. In practice, this is not what we observe. When schools fail, high propensity residents respond more than low propensity. Across all school types the effect size increases as we go up the turnout distribution. For example, when we consider elementary school failure, we see that exposure to treatment raises turnout in the top quantile by 13% (compared to marginally passing schools) while only moving turnout in the bottom quantile about a third as much, increasing turnout by only 5%. These differences are substantively meaningful—the upper coefficient is about 2.6 times the size of the lower—and statistically distinct. This pattern follows in the closest school match and with middle schools, but only slightly when high schools fail.

Figure A.8: Performance Information & Turnout Inequality



Note: Results from quantile regression discontinuity models. Figure plots the effect of failure across the various turnout quantile levels. Unit of analysis is the school-year. Coefficients are based on models with a quartic polynomial of the running variable and the full bandwidth. Standard errors are bootstrapped.

References (Appendix)

Ahn, Thomas and Jacob Vigdor. "The Impact of No Child Left Behind's Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina." *NBER Working Paper* (w20511), (2014a).

Black, Sandra E. "Do better schools matter? Parental valuation of elementary education." *Quarterly journal of economics* (1999): 577-599.

Calonico, Sebastian, Matias D. Cattaneo, and Rocıo Titiunik. "Robust data-driven inference in the regression-discontinuity design." *Stata Journal* 14, no. 4 (2014): 909-946.

Caughey, Devin, and Jasjeet S. Sekhon. "Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008." *Political Analysis* 19, no. 4 (2011): 385-408.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder. "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races." *American Journal of Political Science* 59, no. 1 (2015): 259-274.

Ehrensal, Patricia AL, and Patricia F. First. 2008. "Understanding School Board Politics: Balancing Public Voice and Professional Power." in *Handbook of education politics and policy*. Cooper, Bruce S., James G. Cibulka, and Lance D. Fusarelli, eds. (73-88) Routledge, 2014.

Figlio, David N., and Maurice E. Lucas. "What's in a grade? School report cards and the housing market." *American Economic Review* (2004): 591-604.

Green, Donald P., and Alan S. Gerber. 2008. *Get out the Vote: How to Increase Voter Turnout*. Washington, DC: Brookings Institution Press.

Green, Donald P., Shang E. Ha, and John G. Bullock. "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, no. 1 (2010): 200-208.

Hirschman, Albert O. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Cambridge, MA: Harvard University Press, 1970.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* 105, no. 04 (2011): 765-789.

Imbens, Guido, and Karthik Kalyanaraman. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79, no. 3 (2012).

Jacobsen, Rebecca, Andrew Saultz, and Jeffrey W. Snyder. "Informing or Shaping Public Opinion? The influence of School Accountability Data Format on Perceptions of School Quality" Paper Prepared for The Association for Education Finance and Policy–37th Annual Conference 2012.

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142, no. 2 (2008): 698-714.

Pierson, Paul. "When Effect Becomes Cause: Policy Feedback and Political Change." *World politics* 45, no. 04 (1993): 595-628.

Stock, James H. & Motohiro Yogo. Testing for weak instruments in linear IV regression. In *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. Donald W. K. Andrews, James H. Stock, Thomas J. Rothenberg, eds., 2005.