The effect of Race-Accountability on Student Achievement*

Robert Arons†‡§

February 25, 2020

Abstract

The No Child Left Behind Act's (NCLB) stated goal was to close the achievement gap. I use a regression discontinuity framework to analyze the NCLB Act's attempt to reduce the achievement gap by requiring sociodemographic groups to meet group based targets. I study this question using a panel of North Carolina public school students' test scores in grades three through eight from 2003-2012. I find reading and math scores improve when schools are held accountable for the achievement of each separate group.

Keywords: No Child Left Behind Act, Regression Discontinuity Design

JEL classification: I21, I24

*The views expressed herein are entirely those of the authors and should not be purported to reflect those of the US Department of Justice.

†Email: r.abraham.arons@gmail.com

‡Present Address:US Department of Justice, Antitrust Division, Economic Analysis Group, 450 5th St. NW, Washington DC 20530.

§This work was done at The University of Rochester, 500 Joseph C. Wilson Blvd. Rochester, NY 14627

1 Introduction

The No Child Left Behind (NCLB) Act attempted to close the achievement gap. It did so by requiring certain defined sociodemographic groups of students[1] in North Carolina Title 1 schools that had 40 or more eligible students[2] to meet the same proficiency standards as the school as a whole or the school could face certain types of sanctions. This paper examines whether test scores and proficiency rates for the students in various sociodemographic groups were improved due to this provision. In order to explore this issue, student test score data from North Carolina public schools is utilized.

This paper builds on past work on high-stakes testing. Carnoy and Loeb (2002) found accountability does improve student achievement, and Dee and Jacob (2011) found NCLB increased achievement in math but not reading. In contrast, Reardon et al. (2012) found evidence that gains in minority achievement in the NCLB era were due to pre-existing trends, and Neal and Schanzenbach (2010) that showed high-stakes testing made schools focus on students close to passing. Additionally, Davidson et al. (2015) found the idiosyncratic way states implemented NCLB resulted in school failure rates that differed substantially across states. Worryingly, Sims (2013) found that students performed worse in subsequent years after a sociodemographic group failed. The poor subsequent performance could be evidence that NCLB sanctions harmed students in the schools they were meant to help. By contrast, Ahn and Vigdor (2014) examined the effect of NCLB sanctions using a regression discontinuity framework to compare schools that just met proficiency standards to schools that nearly met them. They found low-performing students experienced improved test scores in subsequent years. Holbein and Ladd (2017) explored changes in school and student behavior for these schools using the same regression discontinuity framework as Ahn and Vigdor (2014)

[1]including asian, black, Hispanic, white, and multi-racial, Native-American, disabled, and limited English proficient

[2]A student is eligible if they have 140 days or more of school membership.

and found schools that students in schools that had faced sanctions from poor performance in past years had fewer unexcused absences, reduced tardiness, and an increase in reported misbehavior.

This paper makes two contributions. First, I evaluate the effectiveness of sociodemographic requirements. I show that these requirements had a positive and significant impact on targeted sociodemographic achievement gaps as measured by test scores and proficiency, they raise scores by about one-tenth of a standard deviation. Second, by applying a regression discontinuity design (RD), I am able to disentangle the effect of the targets for the sociodemographic groups from the other parts of the NCLB that were instituted at the same time, this would not be possible with a traditional difference-in-differences framework. The RD also overcomes the issue that schools with large minority populations may differ in unobservable ways from schools that do not have large minority populations (Kinsler, 2011).

2 Estimation and Identification

The main estimation is done semi-parametrically with local polynomials in the number of students and linear terms in other covariates. Student outcome Y_{it}, which can represent either math or reading proficiency or normalized math or reading test scores, is a function of a non-parametric component of the number of students in the group, $g(Z_{it})$ plus other student and school level covariates which enter linearly, $\beta\mathbf{X_{it}}$. Z_{it} represents the number of students in the sociodemographic group that sat for the test and had at least 140 days of membership. $\mathbf{X_{it}}$ includes year and grade dummies, the grade configuration of the school (1-5, 1-8, 6-8, 5-8), the location of the school, urban, rural, or suburban, the student's last year test score, race dummies (when applicable), percent minority in the school, and quadratics in teacher turnover and lagged test scores. The model is shown in equation 1.

$$Y_{it} = g(Z_{it}) + \beta \mathbf{X_{it}} + \epsilon_{\mathbf{it}} \tag{1}$$

The mainline estimation is done in two steps. In the first step, the estimated residuals $\hat{\epsilon}_{it}$ are obtained from the regression of Y_{it} on $\mathbf{X_{it}}$ and Z_{it}. By construction, $\hat{\epsilon}_{it}$ is orthogonal to $\mathbf{X_{it}}$. In the second step, a local polynomial regression of Y_{it} on Z_{it} is performed. The reason I employed a two-step estimator is that it allows for the addition of control variables, $\mathbf{X_{it}}$, while not suffering from the curse of dimensionality that would occur if the controls were added non-parametrically. Additionally, performing the regression in two steps makes use of the data far away from the threshold to identify β while still allowing non-parametric identification of the treatment using only data near the threshold. A reasonable number of bandwidths are shown so that it is possible to understand how the result varies with bandwidth size. A triangle kernel is used because it has good boundary value properties (Lee and Lemieux, 2010). The estimation was done with 1st, 2nd, 3rd, and 4th degree polynomials, for brevity only 1st and 2nd are shown.

The identification strategy assumes treatment takes place at or above the threshold, 40 students. The estimate of the treatment effect is the difference in the limits of the observed expected outcomes taken above and below the threshold. Holding $\mathbf{X}$ constant, allow Z to approach the threshold from above z^+ and below z^-. The difference between these two expected outcomes is the treatment effect, TE, as in equation 2.

$$TE = \quad limE[Y \mid Z = z^+] - limE[Y \mid Z = z^-] \tag{2}$$

Where the first limit is taken as Z goes to the threshold value, z, from above, and the second limit is taken as Z goes to z from below.

There are several common assumptions used to identify of the treatment effect. First, ϵ_{it}

must vary smoothly with Z_{it}. Most likely this holds because there are no known jumps in the way students are taught or learn when there are 39 versus 40 students in a sociodemographic group. Second, treatment must occur at the threshold. Indeed, administrators have counts of the number of students enrolled in the school very early in the school year; thus, they are likely to know when a sociodemographic group will be above the threshold number of students. Finally, there must be a strictly positive number of students near the threshold. As in many RD frameworks, estimation involves some extrapolation for one of the limits, in this case the lower one, because the number of students cannot come arbitrarily close to 40 because it is a discrete variable. Moreover, I assume assignment to the treatment is as good as random, it is reasonable to assume this because assignment of students to schools is done through a school boundary system, the precise number of students who enter a school is outside of the school's control.

3 Data

This study makes use of both school and anonymized student level data. The student level data was provided by the North Carolina Education Research Data Center. The school level data comes from publicly available administrative data. NCLB mandated testing and sanctions based on performance for third through eighth grade and 10th grade, I only examine third through eighth grade because it is relatively uncommon for secondary schools to have groups near the 40 student cutoff. Students in non-Title 1 schools are not used in the analysis because they not have to implement sanctions and remediation in accordance with NCLB.

The analysis makes use of all students in Title I schools from the school years 2002-2003 to 2011-2012.[3] There are 2,752,511 observations from 1,244 schools and 1,077,616

[3] In May 2012, North Carolina received a flexibility waiver to significantly change the way it implemented NCLB including allowing local school district to decide whether to sanction failing schools.

unique student ids. It appears that discrepancies in the way students were anonymized has created 766 repeats of masked student IDs, these are dropped in the analysis.[4] Both reading and math test scores are normalized so that the test-year-grade mean is zero and standard deviation is one. Students with disabilities or who are limited English proficient are not used in the analysis.

4 Results

The overall evidence is consistent with group accountability having a positive effect on reading and math scores and proficiency. The smoothed scores are plotted against group size in Figure 1. There is a clear jump in math and reading scores at the 40 student cutoff. This discontinuity is about 0.08 standard deviations in reading and math, about one fifth of the overall achievement gap.[5] Similar results hold for the residualized[6] scores, as shown in Figure 2. Proficiency rates exhibit a similar pattern. In Figure 3, math and reading proficiency rates jump up at the threshold by about one percentage point. The same is true of residualized proficiency, shown in Figure 4. Group size and performance seems to follow a similar pattern, regardless of the performance measure. Below the threshold, between 0 and 30 students, proficiency and test scores tend to decrease with group size. There is a jump in performance at 40 students, due to the effect of NCLB. Performance seems to rise after 40 students.

Next, the impact of group based accountability targets on proficiency rates and test scores is analyzed in a non-parametric framework using local polynomials with a triangle kernel.

[4]For more information on how student IDs are created and matched see the technical reports listed in the bibliography (Technical Report #2 and Technical Report #4) available from `www.childandfamilypolicy.duke.edu/research/nc-education-data-center/list-files-variables/`

[5]The estimate of the achievement gap used here is the raw difference in means between black and white students, approximately one half a standard deviation in reading and math scores.

[6]by regressing scores against a set of covariates and plotting the residual by group size

As described in the model section, a variety of polynomials and bandwidths are used to ensure the results are not an artifact of the chosen specification. In Table 2, it can be seen that group accountability has had a positive and significant impact on reading test scores. The test scores are normalized to have a standard deviation of one, so the results in columns one through four suggest that group accountability has raised test scores anywhere between 0.03-0.06 standard deviations. This is roughly one tenth of the achievement gap. These results are highly significant. Columns five and six report the results limiting the analysis to schools with over 200 students and under 200 students to investigate whether larger or smaller schools are more likely to ignore smaller groups of students that do not contribute towards accountability. This does not seem to be a significant difference between the two. Similar results are found for math in table 3. A similar analysis is done for reading proficiency in Table 4 and math proficiency in Table 5. Both tables show group accountability had a positive and significant impact on proficiency. Further robustness checks, show that other RD techniques, including splines and higher order polynomials, produce similar results.

The results should be considered an average treatment effect of this element of the law, but note that one should not consider it an effect completely isolated from the rest of NCLB. For example, schools that had one group above 40 students had to consider how they wished to allocate their scarce resources between students in the group and not in the group and between math and reading as both subjects would count towards school failure. Therefore, students outside the group could potentially lose resources, an effect not considered in this paper. Additionally, NCLB had a multitude of effects on schools, in particular, some schools in the treatment or control group might be facing sanctions from poor student performance in years prior that could affect student outcomes in the analysis year. This is not controlled for in the shown specifications.

One concern with any regression discontinuity design is whether the running variable, in this case group size, is manipulable. Figure 5 addresses these concerns. Figure 5 is the distribution of group size across Title 1 schools in North Carolina. The distribution of group size is relatively continuous across the 40 student threshold. If schools were manipulating group sizes to down to 39, then we would see a large mass of students at 39 and a proportionally smaller number of students at 40 and 41. In fact, it appears that there is a larger mass of students above the 40 student threshold. Although this distributional feature is most likely random variation, if it were due to manipulation, then it would suggest administrators were trying to push their school into accountability for these groups, an unlikely scenario. A test for manipulation defined in McCrary (2008) shows no manipulation. Further robustness checks, available on request, show that other covariates do not jump at the threshold.

5 Conclusion

I conclude that group accountability does have a statistically significant positive effect on the test scores and proficiency rates of groups. Group test scores improved about 0.03-0.06 standard deviations because of group accountability. This is about one-tenth of the achievement gap.

Table 1: Descriptive Statistics

	Mean	SD
Normalized math score	0	1
Normalized reading score	0	1
Proficient in math	0.76	
Proficient in reading	0.70	
Asian	0.019	
Black	0.32	
Hispanic	0.12	
Native American	0.020	
Multi-racial	0.034	
Teacher turnover this year	0.16	0.096
Free lunch	0.53	
Days membership	166	4.56
Grades School Serves		
Grades 1-5	0.57	
Grades 5-8	0.025	
Grades 6-8	0.17	
Grades 1-8	0.082	
Grades 1-6	0.11	
Other Grade Config	0.039	

Selected variables and statistics from the individual student records from Title 1 schools in North Carolina, school years 2002-2003 to 2011-2012. Standard deviation not shown for binary variables.

Table 2: Estimation with Local Polynomials: Reading Score

	(1)	(2)	(3)	(4)	(5)	(6)
Group $\geq$ 40	0.06**	0.03***	0.06**	0.04***	0.03***	0.02**
	0.03	0.01	0.03	0.02	0.01	0.01
Other covariates	No	No	No	Yes	No	No
Polynomial degree	1	2	1	1	1	1
Bandwidth	5	5	3	5	5	5
Schools over 200 students	No	No	No	No	Yes	No
Schools under 200 students	No	No	No	No	No	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Data comes from the administrative records of North Carolina 3-8th graders in Title 1 public schools from school years 2002-2003 to 2011-2012. Covariates include dummies for grade configuration, ethnicity, free lunch status, school location (urban, rural, etc), year, as well as the percent minority, quadratics in teacher turnover, lagged reading and math scores. The local polynomials are taken in the number of students in the subgroup. Robust clustered standard errors are at the school year level.

Table 3: Estimation with Local Polynomials: Math Score

	(1)	(2)	(3)	(4)	(5)	(6)
Group $\geq$ 40	0.04***	0.03**	0.04***	0.02	0.03***	0.02**
	0.01	0.01	0.01	0.02	0.01	0.01
Other covariates	No	No	No	Yes	No	No
Polynomial degree	1	2	1	1	1	1
Bandwidth	5	5	3	5	5	5
Schools over 200 students	No	No	No	No	Yes	No
Schools under 200 students	No	No	No	No	No	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Data comes from the administrative records of North Carolina 3-8th graders in Title 1 public schools from school years 2002-2003 to 2011-2012. Covariates include dummies for grade configuration, ethnicity, free lunch status, school location (urban, rural, etc), year, as well as the percent minority, quadratics in teacher turnover, lagged reading and math scores. The local polynomials are taken in the number of students in the subgroup. Robust clustered standard errors are at the school year level.

Table 4: Estimation with Local Polynomials: Reading Proficiency

	(1)	(2)	(3)	(4)	(5)	(6)
Group $\geq$ 40	0.02**	0.02**	0.03***	0.01*	0.02**	0.02**
	0.01	0.01	0.01	0.01	0.01	0.01
Other covariates	No	No	No	Yes	No	No
Polynomial degree	1	2	1	1	1	1
Bandwidth	5	5	3	5	5	5
Schools over 200 students	No	No	No	Yes	Yes	No
Schools under 200 students	No	No	No	No	No	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Data comes from the administrative records of North Carolina 3-8th graders in Title 1 public schools from school years 2002-2003 to 2011-2012. Covariates include dummies for grade configuration, ethnicity, free lunch status, school location (urban, rural, etc), year, as well as the percent minority, quadratics in teacher turnover, lagged reading and math scores. The local polynomials are taken in the number of students in the subgroup. Robust clustered standard errors are at the school year level.

Table 5: Estimation with Local Polynomials: Math Proficiency

	(1)	(2)	(3)	(4)	(5)	(6)
Group $\geq$ 40	0.02***	0.03***	0.03***	0.01*	0.03**	0.01*
	0.01	0.01	0.01	0.01	0.01	0.01
Other covariates	Yes	Yes	Yes	No	Yes	Yes
Polynomial degree	1	2	1	1	1	1
Bandwidth	5	5	3	5	5	5
Schools over 200 students	No	No	No	No	Yes	No
Schools under 200 students	No	No	No	No	No	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Data comes from the administrative records of North Carolina 3-8th graders in Title 1 public schools from school years 2002-2003 to 2011-2012. Covariates include dummies for grade configuration, ethnicity, free lunch status, school location (urban, rural, etc), year, as well as the percent minority, quadratics in teacher turnover, lagged reading and math scores. The local polynomials are taken in the number of students in the subgroup. Robust clustered standard errors are at the school year level.

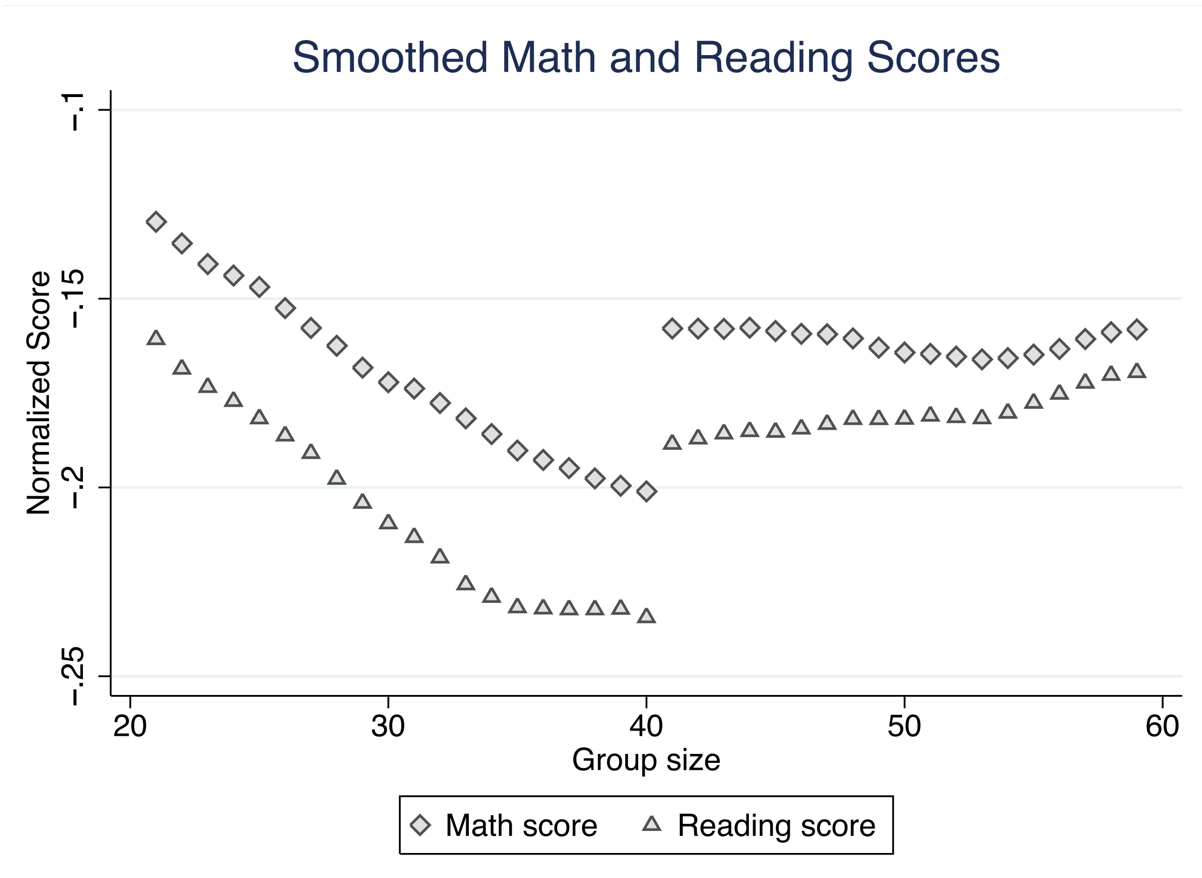

Figure 1: Normalized smoothed math and reading scores come from administrative records of all students in Title I schools from the school years 2002-2003 to 2011-2012. Smoothing is done with zero degree local polynomials with a bandwidth of five using only data from the side of the 40 student cut-off the grid point is on.

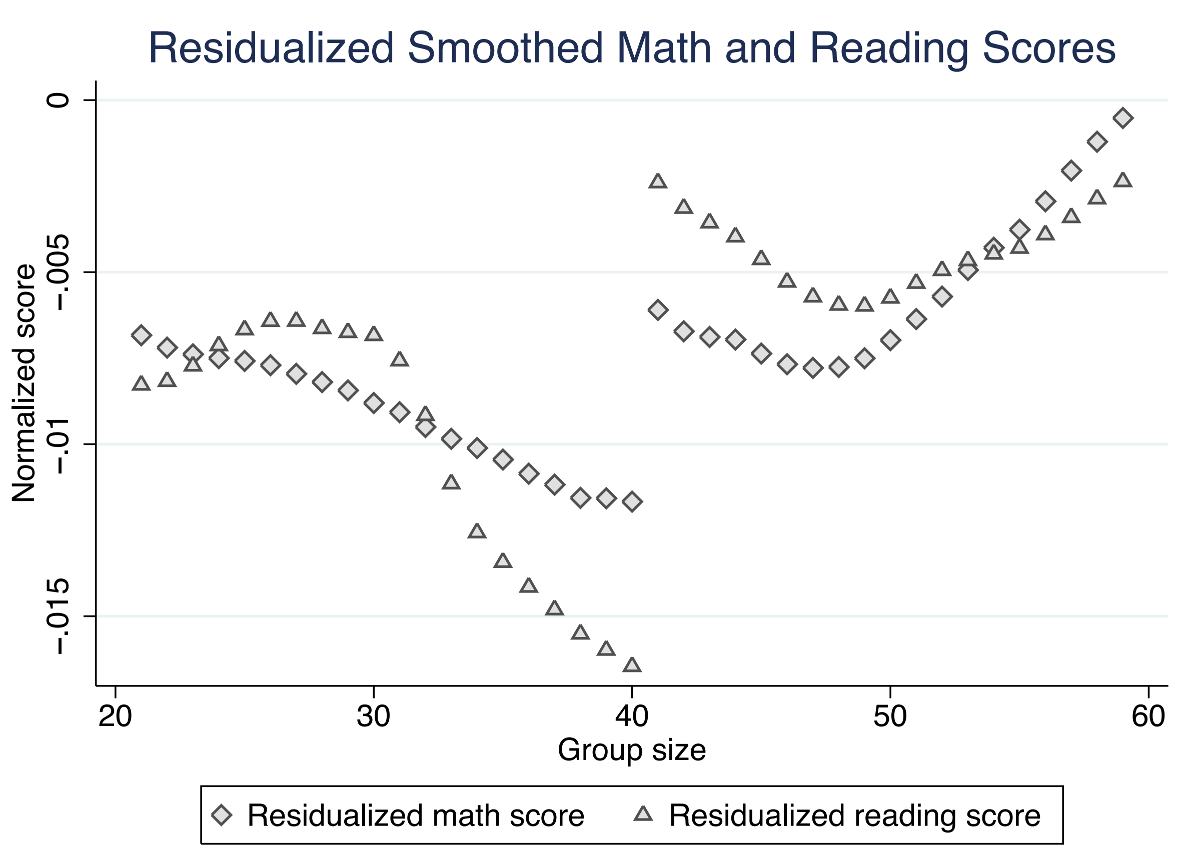

Figure 2: Residualized smoothed math and reading scores come from administrative records of all students in Title I schools from the school years 2002-2003 to 2011-2012. Smoothing is done with zero degree local polynomials with a bandwidth of five using only data from the side of the 40 student cut-off the grid point is on. Test scores are normalized by test-year-grade to have a zero mean and standard deviation of one. Scores are residualized using year and grade dummies, the grade configuration of the school (1-5, 1-6, 1-8, 6-8, 5-8), the location of the school, urban, rural, suburban etc., the student's last year test score, race dummies (when applicable), percent minority in the school, and quadratics in teacher turnover and lagged test scores.

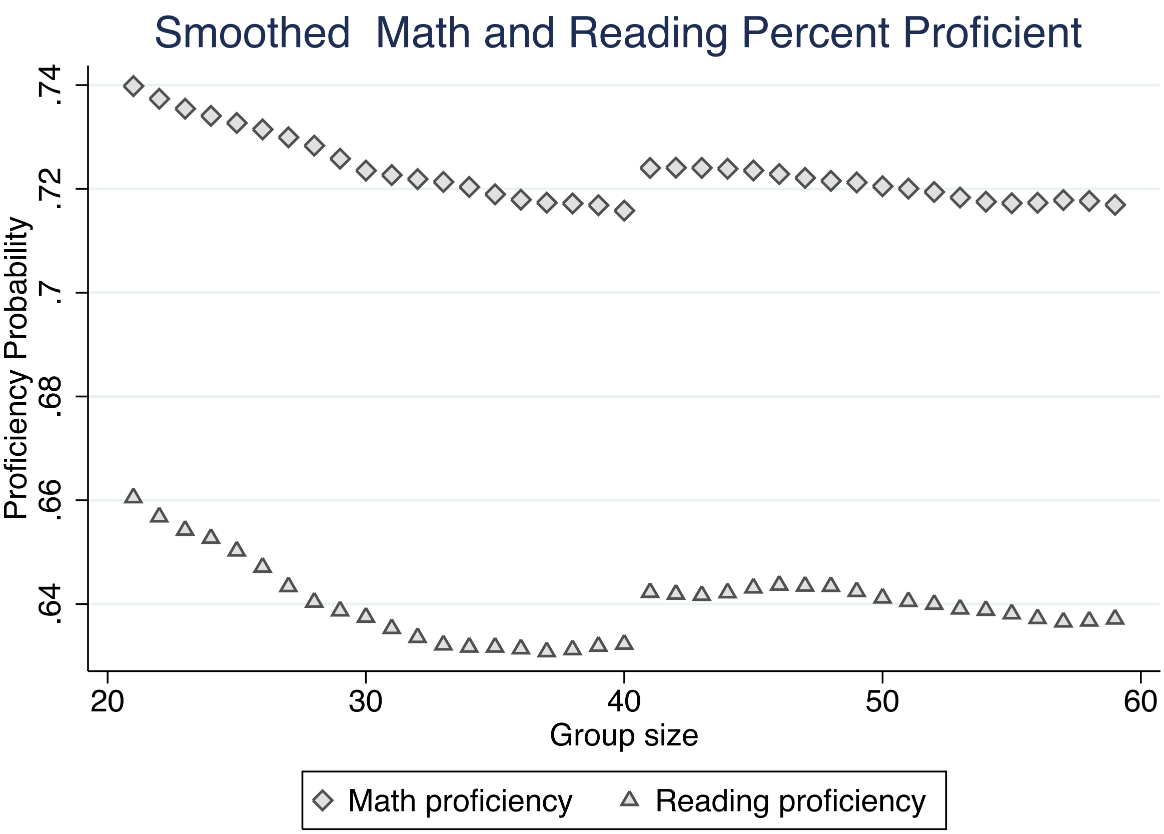

Figure 3: Smoothed math and reading proficiency come from administrative records of all students in Title I schools from the school years 2002-2003 to 2011-2012. Smoothing is done with zero degree local polynomials with a bandwidth of five using only data from the side of the 40 student cut-off the grid point is on.

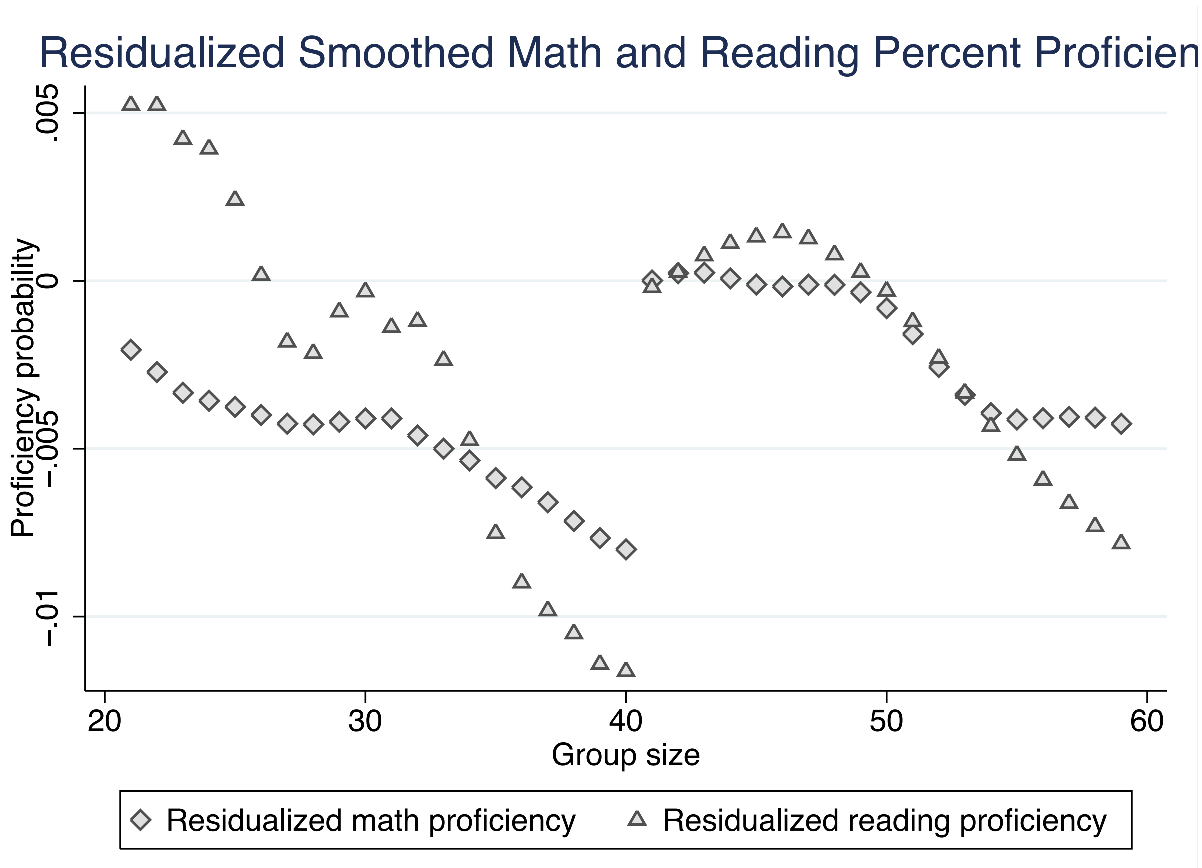

Figure 4: Residualized smoothed math and reading proficiency come from administrative records of all students in Title I schools from the school years 2002-2003 to 2011-2012. Smoothing is done with zero degree local polynomials with a bandwidth of five using only data from the side of the 40 student cut-off the grid point is on. Proficiency indicators are residualized using year and grade dummies, the grade configuration of the school (1-5, 1-6, 1-8, 6-8, 5-8), the location of the school, urban, rural, suburban etc., the student's last year test score, race dummies (when applicable), percent minority in the school, and quadratics in teacher turnover and lagged test scores.

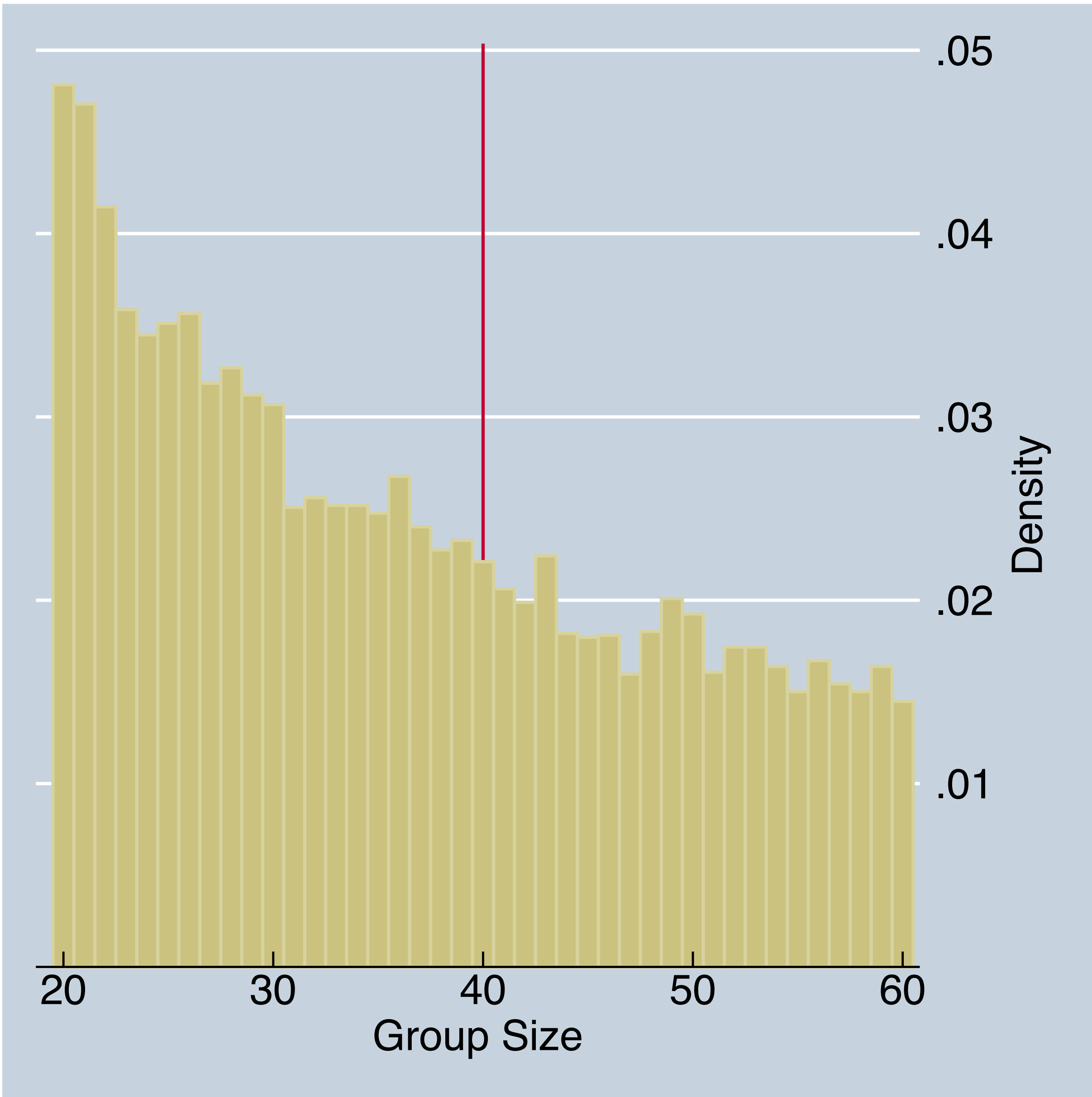

Figure 5: Distribution of Group Size Conditional on Group Size between 20 and 60, All Groups.

References

Ahn, Thomas and Jacob Vigdor, "The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina," Technical Report, National Bureau of Economic Research 2014.

Carnoy, Martin and Susanna Loeb, "Does external accountability affect student outcomes? A cross-state analysis," *Educational Evaluation and Policy Analysis*, 2002, *24* (4), 305–331.

Davidson, Elizabeth, Randall Reback, Jonah Rockoff, and Heather L Schwartz, "Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB," *Educational Researcher*, 2015, *44* (6), 347–358.

Dee, Thomas S and Brian Jacob, "The impact of No Child Left Behind on student achievement," *Journal of Policy Analysis and Management*, 2011, *30* (3), 418–446.

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*, 2017, *58*, 55–67.

Kinsler, Josh, "Understanding the black–white school discipline gap," *Economics of Education Review*, 2011, *30* (6), 1370–1383.

Lee, David S and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *The Journal of Economic Literature*, 2010, *48* (2), 281–355.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Neal, Derek and Diane Whitmore Schanzenbach, "Left behind by design: Proficiency counts and test-based accountability," *The Review of Economics and Statistics*, 2010, *92* (2), 263–283.

Reardon, Sean F, Erica Greenberg, Demetra Kalogrides, Kenneth A Shores, and Rachel A Valentino, "Trends in Academic Achievement Gaps in the Era of No Child Left Behind.," *Society for Research on Educational Effectiveness*, 2012.

Sims, David P, "Can failure succeed? Using racial subgroup rules to analyze the effect of school accountability failure on student performance," *Economics of Education Review*, 2013, *32*, 262–274.