Does No Child Left Behind Have Teeth? Examining the Impact of Federal Accountability Sanctions in North Carolina*

Tom Ahn† Jacob Vigdor‡

October 19, 2009

1 Introduction

School accountability programs operate under the premise that public schools have insufficient incentives to use resources efficiently. The positive or negative sanctions incorporated into the program are intended to provide these incentives. This paper evaluates the impact of incentives associated with the Federal No Child Left Behind (NCLB) Act of 2001 on school performance in North Carolina Public Schools. The effect of threatened sanctions, or of sanctions themselves, is inferred using regression discontinuity (RD) methods. We find that the NCLB sanctions have had no significant impact on student performance in public schools. While the theoretical rationale for incentivizing schools is straightforward, previous literature has provided only a small amount of evidence that students learn more when schools are subject to incentives. There is some cross-state evidence on the impacts of state-level accountability programs (Carnoy and Loeb 2002; Hanushek and Raymond 2002), but state-level accountability initiatives may correlate with other education policies that independently influence achievement. There is also some evidence of impacts associated with programs implemented by individual states or districts (Figlio and Rouse 2006, Peterson and West 2006, Chakrabarti

*Preliminary Draft. Please do not cite.

†University of Kentucky

‡Duke University and NBER

2006), but evaluations of such programs are often reduced-form in nature and thus provide little insight as to the causal mechanisms underlying the effects. Finally, there has been a heated debate regarding the potential for traditional market competition to improve the quality and efficiency of public schools (Hoxby 2000 etc). Previous research has identified a number of perverse consequences of incentivizing schools, generally rooted in the use of standardized test scores to measure a school's output. Schools have been shown to classify marginal students as disabled to prevent their scores from being counted (Cullen and Reback, 2002, Figlio and Getzler 2002, Jacob, 2005), to suspend low-achieving students on test days (Figlio, 2006), and teachers have been shown to alter students' answer sheets manually (Jacob and Levitt, 2003). Other, less blatant attempts to 'game the system' by teaching to the test or focusing on a specific subset of students has been observed as well. (Grissmer and Flanagan, 1998, Neal and Schanzenbach 2009).

The common issue in these papers is that accountability systems are so complex that it is difficult to settle on a focal point. Studies looking at perverse incentives point out interesting behavioral response of teachers and administrators, but are so specific that robust policy recommendations do not usually emerge. On the other side, cross-state analyses are so general that not much more than a proposal for 'stronger' accountability can be made.

There are essentially two components of accountability systems that the policy maker can adjust to make the system more or less effective. One of these is setting the standard that the school must meet in order to qualify for the reward or avoid the punishment. Policy makers must select the correct difficulty for the standard such that it is neither too easy nor too difficult to reach. This ensures that schools must exert additional effort or implement some change for the school to make the grade (See Ahn 2009 for an example.). The second component is the generosity of the reward or the severity of the punishment. A very generous reward or a very severe punishment may spur higher achievement. However, a worthless incentive or a toothless punishment will not elicit any change in behavior or increase education production in response.

In this paper, we attempt to use the strength of the sanction as the single focal point by which the effectiveness of NCLB can be analyzed. The central argument is that it is too difficult to assess the myriad of measures that teachers and principals take (or do not take) to respond to accountability systems. From changing the peer grouping or teacher exposure

of students to installing different didactic methodology in the classroom, a school may try any number of things to increase achievement. In addition, many effective changes may be unobservable to the econometrician, contributing to measurement problems.

On the other hand, measuring the impact of the sanction (or the threat of being sanctioned) captures the most important dimension of the accountability system: its effectiveness. If the cost of being sanctioned is high, teachers and administrators will take steps to ensure that they are not impacted. On the other hand, the idea that NCLB would have changed all schools behavior because they feared the sanctions so much does not make too much sense if all the sanctions are themselves toothless.

Our evaluation of NCLB takes advantage of the fact that sanctions are applied according to whether schools meet specific thresholds in terms of the proportion of students deemed "proficient" according to state standards. There are two forms of exemptions applied to these thresholds, but both effectively serve to move the threshold in a predictable way. We infer the impact of a sanction, or the threat of a sanction, by comparing schools barely exposed to the sanction or threat with those who barely escaped the sanction or exposure.

Our analysis also takes advantage of specific elements of NCLB. Schools must fail to make AYP for two consecutive years before being exposed to any sanctions, which implies that we can make independent inferences about the impact of sanctions themselves and the threat of being exposed to them. We also take advantage of a pilot program implemented in some North Carolina districts that switched the order of sanctions applied. This permits us to separately identify the impact of the first sanction from the impact of being at risk for the second.

In theory, accountability programs can have systemic effects that our RD methodology would not capture. All schools may alter their behavior upon implementation of the program, regardless of whether they are exposed to a specific sanction or threat. We argue, however, that evidence that schools make no adjustments when faced with the threat of a specific sanction strongly suggests that schools also make no adjustments in initial periods when they are not faced with specific sanction threats. Our analysis reveals significant effects neither of sanctions themselves or the threat of being exposed to them.

2 Education Production and NCLB

The canonical economic model of education equates schools with firms. These firms use a combination of inputs, including students, teachers, and instructional materials, to produce human capital. In a competitive market with homogeneous students and firms, we would expect all schools to charge the same price and use the same production technology to produce educated students. Schools that used inputs inefficiently would be unable to offer the same combination of price and quality of service provided, and would be competed out of the market.

The main departure from the competitive marketplace occurs because a specific subset of schools is eligible for government subsidies, which enable them to charge a price of zero for their product. In such a scenario, a school using inefficient production technology could persist, if it continues to receive a subsidy.

Figure 1 shows a stylized version of this scenario, where education relies on a single input and produces a unidimensional output. The production function shows the maximum amount of output obtainable with a given level of input. The point A represents an inefficient outcome, where a school could have obtained a higher level of output given the input level.

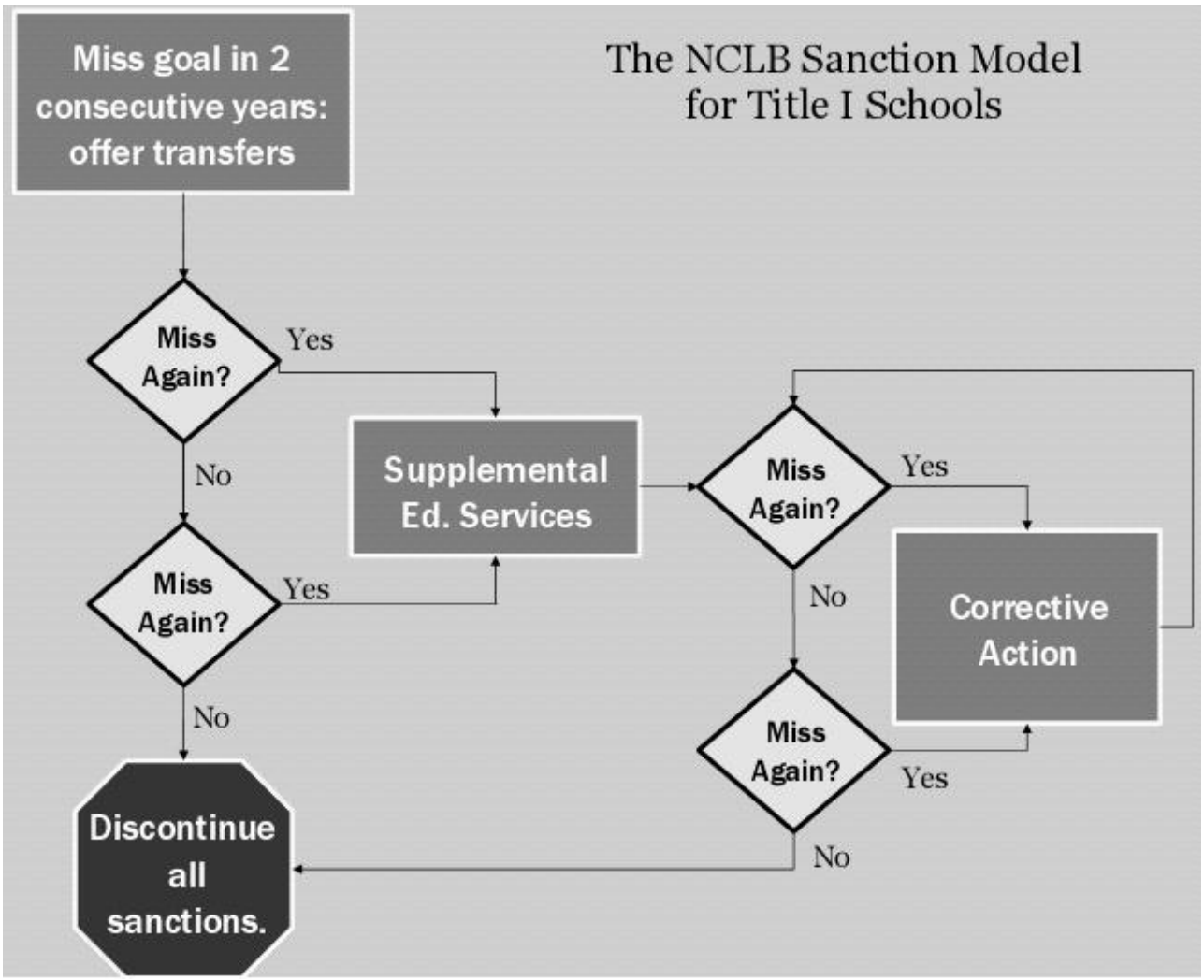
The stated purpose of accountability systems is to mimic the competitive forces that would ordinarily force firms to use efficient production technologies. Whereas inefficiently producing firms in the private sector face bankruptcy, inefficiently producing schools face sanctions in accountability programs. While these sanctions are unlikely to be as costly to a school as bankruptcy to a firm, the hope is that they will move schools in the right direction: toward obtaining more output without a significant increase in inputs. For example, one of the NCLB sanctions attempt to introduce direct market pressure and competition by allowing students in low-performing schools the option to transfer to other public schools within the same district.

3 NCLB and Implementation in North Carolina

North Carolina students in grades 3 through 8 must take end-of-grade exams in reading and mathematics. The test is on a developmental scale, allowing comparison of scores from consecutive grades[1]

[1]High school students must take end-of-course exams.

Figure 1: NCLB Sanction Model in NC



In the NCLB system, which in North Carolina was implemented in addition to the state's pre-existing system[2], schools are subject to sanctions when the proportion of students scoring above a state-specific proficiency threshold, or the proportion of students in any of several demographic, socioeconomic, or linguistic subgroups scoring above the same threshold, falls below mandated levels. When the percent of students scoring at the proficiency level exceeds the threshold in each measured subgroup and the school as a whole, the school is said to have made "adequate yearly progress" (AYP). Schools that fail to meet the AYP standard face sanctions or the threat of sanctions, depending on their past history. The NCLB sanctions apply at the school level rather than the individual teacher level.

North Carolina's standardized tests collapse scores into four ordinal categories, the second highest of which is proficient. Schools make adequate yearly progress when a pre-determined fraction of all students, and an identical fraction of students in each of a set of subgroups score at the proficient level or above. The performance of subgroups is tracked so long as there are at least 40 tested students in the subgroup. Subgroups include racial minorities, the economically disadvantaged, the disabled, and students classified as having limited English proficiency. The school must also maintain a 95 % attendance rate.

Failure to meet the AYP standard in two consecutive years places a school in the sanction regime. Schools remain in this regime until they meet the AYP standard in two consecutive years. In addition to satisfying the state requirement as outlined above, there are two important methods by which schools can be considered to make AYP (and therefore be exempt from sanctions) even if they fail to meet the requirement.

The first exemption is called safe harbor. A school that failed to make AYP may be categorized as passing under the safe harbor condition if the subgroup that results in the school being classified as not making AYP made at least a 10 % improvement in the proficiency rate. A school may also make AYP under an exemption called confidence interval. In essence, the state draws a 95 % confidence interval (CI) around the actual performance of the school, and if the interval contains the state mandated proficiency rate, the school is categorized as making AYP.

The first sanction, which applies to schools that miss AYP in the same subject for two consecutive years, is to offer students space-available transfers to higher-performing public schools

[2]For a complete description of the state's separate accountability system, refer to Vigdor (2007).

in the same district. The second, which applies to schools that miss AYP one or two years after missing it twice in a row, involves offering "supplementary education services" – tutoring - to selected students. A fourth failure to achieve the AYP standard results in "corrective action," which can mean any of several administrative moves ranging from extending the school year to replacing some staff. Not shown in the diagram is a final stage in the sanction regime, restructuring, which applies after the fifth failure to meet AYP. Restructuring could involve turning over operation of the school to another entity, replacing the staff, or other measures. Schools can exit the sanction regime by making AYP for two consecutive years.

Analyzing accountability systems in North Carolina presents an additional problem because of the existence of the state specific ABC accountability system, which offers cash incentives for student achievement gains. NCLB was layered on top of the ABC system, using the same end-of-grade examination as the metric by which school performance is judged. Besides the difference in incentives structure, while NCLB judges school success by proficiency rates, ABC judges success by whether the average test score increase in the school is greater than the state average observed between 1992/93 and 1993/94, the first two years in which the test was administered.

Because the two accountability systems are so different, the two systems are not highly correlated and sometimes give conflicting signals about the performance of schools. In many cases a school will be judged to be qualifying under one system, yet failing according to the other system. Because of the complex interaction between the two accountability systems, controlling for ABC status in a conventional regression is unsatisfactory. In particular, a school could be attempting to satisfy NCLB and ABC jointly, or it could be ignoring one (or both) of the standards. A school may choose to devote different resources in differing intensities, depending on how close the school is to one standard compared to the other. The usual average treatment effect interpretation is undesirable for these reasons. We implement a RD framework, precisely because the internal validity characteristics of the RD estimator allow us to analyze only the impact of the NCLB sanction.

4 Data

We use an administrative dataset for the North Carolina public school system from the academic years 2002/03 to 2005/06.[3] The dataset contains information on all public schools, students, and teachers in North Carolina. The data is collected annually, and the capability exists to link students and teachers across years. What emerges is a relatively complete longitudinal picture of the entire North Carolina public school system.

For the purposes of this paper, we aggregate the data to the school level. In essence, what we are interested in is the school's response to the threat of NCLB sanctions, and how this threat translates into average academic achievement. Table 1 summarizes the school characteristics and performance measures of the data set for all schools and schools separated by whether they make AYP or not.

As one may expect, schools that do not make AYP in general have a higher percentage of traditionally under-privileged racial and socio-economic subgroups. Students in these failing schools have lower attendance and promotion rates and are taught by less experienced teachers. Unsurprisingly, various outcome measures are all higher for schools that make AYP.

5 Empirical Framework

In this section, we specify the empirical framework. For each school j in year t, let $W_{jt} \in \{0, 1\}$ define the treatment received, where $W_{jt} = 1$ indicates that the school received the treatment in the current year and $W_{jt} = 0$ otherwise. In our study, the treatment is defines as the threat of NCLB sanctions. That is, there is actually no set 'program' that the school take part in when $W_{jt} = 1$. The question at hand is, given that the school, due to its poor performance, faces the possibility of sanctions described in the previous section, does it take steps to make AYP in the following year, thus avoiding sanctions?

Let $Y_{j,t+1}(1)$ be some outcome measure of interest for school j in the following year if it is currently under threat of sanctions, and let $Y_j(0)$ be the outcome next year if the school

[3]The data, which is collected by NCDPI, was made available by North Carolina Education Research Data Center (NCERDC: www.pubpol.duke.edu/centers/child/nceddatacenter.html) at the Center for Child and Family Policy. While student and teacher level data are confidential, aggregate data and summary statistics are publicly available at the NCDPI web site (www.ncpublicschools.org/reportstats.html).

Table 1: Summary Statistics

	All Schools	Made AYP	Failed AYP
School-wide Proficiency year t	0.8339 (0.0872)	0.8622 (0.0703)	0.7748 (0.0928)
School-wide Proficiency year t+1	0.8371 (0.0858)	0.8599 (0.0749)	0.7895 (0.0911)
Worst Group Proficiency year t	0.6725 (0.1743)	0.7667 (0.1129)	0.4894 (0.1352)
Worst Group Proficiency year t+1	0.6889 (0.1696)	0.7487 (0.1396)	0.5535 (0.1615)
% minority	0.3851 (0.2780)	0.3284 (0.2640)	0.5034 (0.2757)
% free-lunch	0.4971 (0.2264)	0.4595 (0.2214)	0.5787 (0.2220)
% daily attendance	0.9560 (0.0080)	0.9566 (0.0079)	0.9546 (0.0081)
% 3rd grade promoted	0.9700 (0.0382)	0.9738 (0.0328)	0.9611 (0.0486)
% 5th grade promoted	0.9887 (0.0789)	0.9900 (0.0216)	0.9867 (0.0287)
% teachers 4 - 10 years experience	0.2743 (0.0886)	0.2739 (0.0905)	0.2760 (0.0855)
% teachers $>$ 10 years experience	0.4945 (0.1318)	0.5088 (0.1321)	0.4646 (0.1292)
Made AYP	0.7002 (0.4582)		
transfer sanction	0.1927 (0.3944)	0.0023 (0.0474)	0.6329 (0.4822)
tutoring sanction	0.0987 (0.2983)	0.0154 (0.1233)	0.2873 (0.4527)
Observations	4,194	2,658	1,114

is not under threat. Then, the outcome for school j in year $t+1$ is defined as: $Y_{j,t+1} = W_j Y_{j,t+1}(1) + (1 - W_j)Y_{j,t+1}(0)$. This can be re-written as $Y_{j,t+1} = \beta + \alpha_{jt} W_{jt} + \epsilon_{jt}$ where $\alpha_{jt} = Y_{j,t+1}(1) - Y_{j,t+1}(0)$. In this sense, $E(\alpha_{jt})$ is the treatment effect we wish to estimate. Since we can never observe a school that is both treated and untreated at the same time, differencing the average outcome for schools that are under threat of sanctions and schools that make AYP will not yield the desired treatment effect. In particular:

$$
\begin{aligned}
& E(Y_{j,t+1}|W_j = 1) - E(Y_{j,t+1}|W_j = 0) \\
= \; & E(\alpha_{jt}) + (E(Yj, t+1(0)|W_{jt} = 1) - E(Yj, t+1(0)|W_{jt} = 0)) + Pr(W_{jt} = 0) \cdot \\
& \quad (E(Y_{j,t+1}(1) - Y_{j,t+1}(0)|W_{jt} = 1) - E(Y_{j,t+1}(1) - Y_{j,t+1}(0)|W_{jt} = 1) \\
& \quad\quad -E(Y_{j,t+1}(1) - Y_{j,t+1}(0)|W_{jt} = 1) - E(Y_{j,t+1}(1) - Y_{j,t+1}(0)|W_{jt} = 0)) \\
= \; & E(\alpha_{jt}) + (E(\epsilon_{jt}|W_{jt} = 1) - E(\epsilon_{jt}|W_{jt} = 0)) + Pr(W_{jt} = 0)(E(\alpha_{jt}|W_{jt} = 1) - E(\alpha_{jt}|W_{jt} = 0)
\end{aligned}
$$

The difference between the average outcomes for the two groups of schools will not equal the treatment effect unless the second and third terms in the last line of the equation equals zero. There are several reasons to believe that this will not be the case. First and foremost is the complication of the ABC accountability system briefly described above. Because ABC is not perfectly correlated with NCLB, it is not redundant in the system (allowing us to treat the impact as a joint NCLB and ABC effects.). At the same time, because both accountability systems use the end-of-grade exam as the measure of achievement, high performance under one system may imply high performance under the other system. In this way the two systems must be correlated to some degree. In addition, when schools are threatened with sanctions, we may expect them to take any variety of steps to improve performance. However, schools that performed well enough to make AYP may already have been implementing some of these "treatment" measures already.

To overcome these issues, we use the fact that NCLB is implemented in such a way that schools that just make AYP face no sanction pressures while observationally equivalent schools that just fail to make AYP face sanction pressures. Using a RD approach, we can exploit this structural break to discern if NCLB sanctions actually put pressure on schools to perform better.

5.1 Constructing the Treatment Criterion

Let x_j be a continuous variable that partially determines whether school j receives the treatment. In our framework, this is the relative proficiency rate of the lowest performing subgroup, as defined below. First, we calculate the proficiency rate of all relevant subgroups in the school and select the subgroups that fall below the state criterion. We then apply the safe harbor and CI exemptions as described above. If all subgroups pass under the exemption rules, the school is considered making AYP. Therefore, defining c as the state proficiency rate that each subgroup in the school must beat, proficiency rate of subgroup k in school j is high enough to make AYP in year t if:

$$x_{jkt} \geq c_{jkt} \qquad \forall \qquad k \in [1, K]$$

where:

$$c_{jkt} = \min\{x_{jk,t-1} * 1.1, \underline{c}, \underline{c} - \delta\}$$

Here, $\underline{c}$ is the state-wide proficiency rate that is usually advertised as the requirement for each subgroup, and δ is the 95% CI. We define $\min\{x_{j1t} - c_{j1t}, x_{j2t} - c_{j2t}, \ldots, x_{jKt} - c_{jKt}\}$ as the relative proficiency score of the lowest performing subgroup. We note that the subgroup k is not a mutually exclusive grouping. That is, students may be located in multiple subgroups, such as a disabled Hispanic student who has limited English proficiency (3 subgroups).

Notice that the subgroup specific criterion is not really a minimum of the three values (last year performance + 10 %, state requirement, and state requirement - CI), as CI measure is always smaller than just the state requirement. Therefore, any criterion that fails to account for safe harbor and CI calculation will not result in a true treatment variable. It is interesting to view the differences in success in predicting treatment when we do and do not account for safe harbor and CI. Figure 2 and Figure 3 demonstrate our ability to predict whether a school is threatened with sanctions based on our observation of the school's performance and our understanding of the accountability rules in place. The horizontal axis represents the relative proficiency score of the lowest performing subgroup, as defined above. The vertical axis is the proportion of schools with the same proficiency score profile that made AYP, according to the data.

Table 2: Fit without Safe Harbor and Confidence Intervals

AYP Target	School met AYP		
Met Index	No	Yes	Total
No	1,238	19	1,257
Yes	1,227	1,710	2,937
Total	2,465	1,729	4,194

Figure 2 shows that if we do not account for the exemption policies, we would not be able to predict whether a school with a particular proficiency rate profile will receive the treatment. There is no discontinuity where the relative proficiency score equals zero, which should be the bar set by the state for AYP status. Table 2 shows that we do not do a very good job of predicting from the data, whether a school made AYP or not. Once we account for safe harbor and CI, a discontinuity appears where the score equals zero. This is noticeable in Figure 3. Schools to the left of the criterion have subgroups where $\min\{x_{j1t} - c_{j1t}, x_{j2t} - c_{j2t}, \ldots, x_{jKt} - c_{jKt}\} < 0$. Schools to the right of the criterion have subgroups such that $\min\{x_{j1t} - c_{j1t}, x_{j2t} - c_{j2t}, \ldots, x_{jKt} - c_{jKt}\} > 0$. As such, schools to the right of the zero line are much more likely to make AYP while schools to the left are unlikely to make AYP.[4] Table 3 shows that we do a much better job of predicting whether a school made AYP after accounting for the exemptions.

As a discontinuity exists at the proficiency score of zero, we estimate the impact of sanctions using a fuzzy RD framework (See Imbens and Lemieux 2009.). We use local linear regression to estimate parameters for the outcome and the treatment indicator on both sides of the discontinuity point, and calculate the ratio of the two discontinuities. First, the local linear

[4]To note is that the discontinuity is not 'sharp.' Technically, all schools to the left of the criterion should have no chance at making AYP, and all schools to the right of the criterion should have a 100 % chance of making AYP. The 'fuzzy' discontinuity can be from many sources. One possibility is measurement error. Another is that we have not completely captured the rule set for AYP status determination.

Table 3: Fit including Safe Harbor and Confidence Intervals

AYP Target	School met AYP		
Met Index	No	Yes	Total
No	1,114	143	1,257
Yes	279	2,658	2,937
Total	1,393	2,801	4,194

Figure 2: Treatment Variable without Exemptions

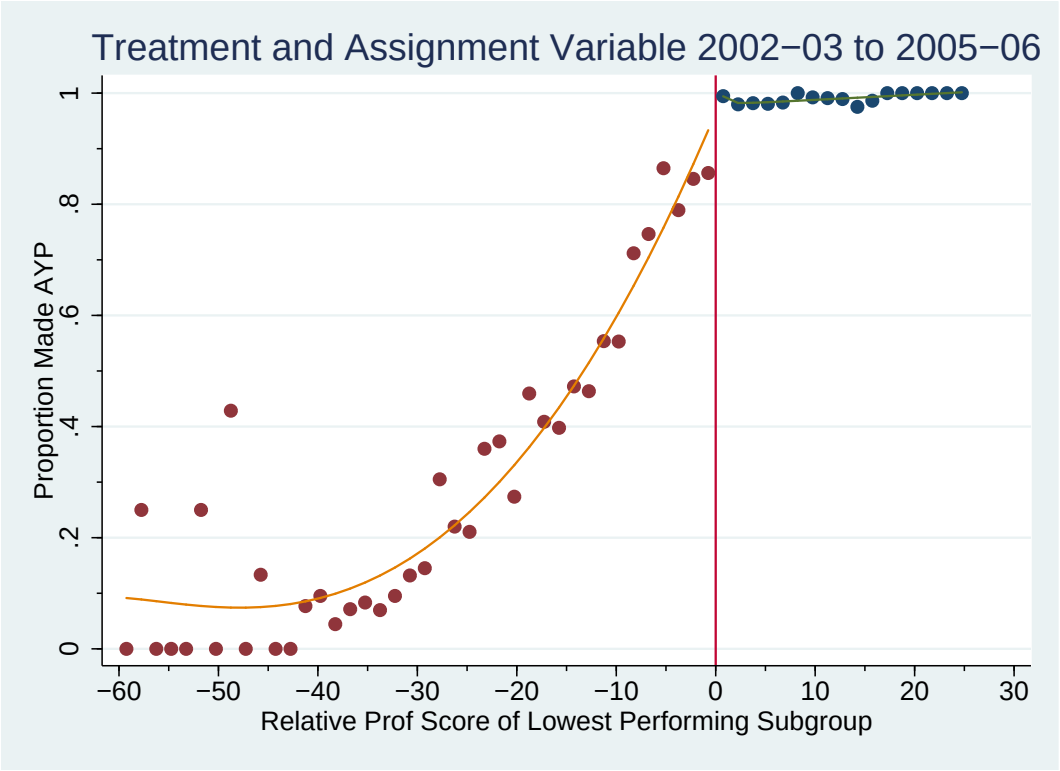
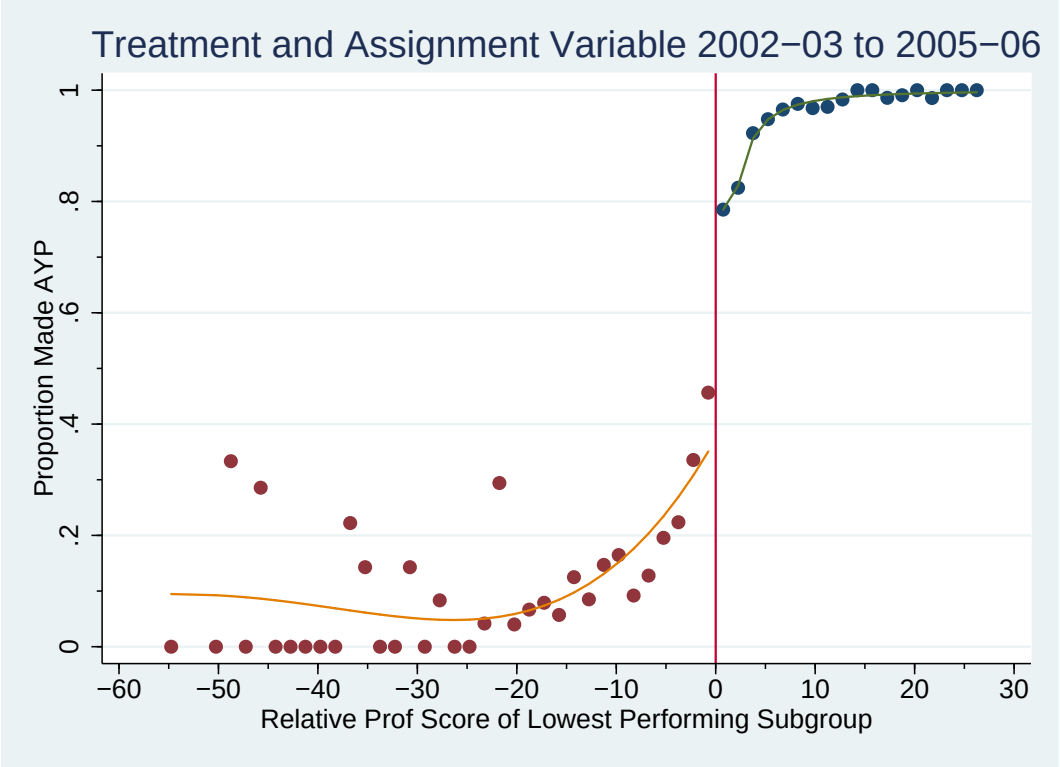


Figure 3: Treatment Variable with Exemptions



regression for the outcome variable on both sides of the discontinuity is:

$$(\widehat{\alpha_{yl}}, \widehat{\beta_{yl}}) = \arg \min_{\alpha_{yl}, \beta yl} \sum_{jt: -h \leq pr_{jt} < 0} (Y_{j,t+1} - \alpha_{yl} - \beta_{yl} \cdot pr_{jt})^2 \quad (1)$$

$$(\widehat{\alpha_{yr}}, \widehat{\beta_{yr}}) = \arg \min_{\alpha_{yr}, \beta yr} \sum_{jt: 0 \leq pr_{jt} \leq h} (Y_{j,t+1} - \alpha_{yr} - \beta_{yr} \cdot pr_{jt})^2 \quad (2)$$

where $pr_{jt} = \min\{x_{j1t} - c_{j1t}, x_{j2t} - c_{j2t}, \ldots, x_{jKt} - c_{jKt}\}$ is defined as the relative proficiency score of the lowest performing subgroup at school j. The binwidth h is chosen to have at least 30 observations on either side of the discontinuity when calculating the conditional mean. There are several outcome variables of interest. For instance, we may be interested in the proficiency rate of the school in year $t+1$ (a level effect) or change in the school average score (a growth effect). Second, the local linear regression for the treatment variable on both sides of the discontinuity is:

$$(\widehat{\alpha_{wl}}, \widehat{\beta_{wl}}) = \arg \min_{\alpha_{wl}, \beta wl} \sum_{jt: -h \leq pr_{jt} < 0} (W_{jt} - \alpha_{wl} - \beta_{wl} \cdot pr_{jt})^2 \quad (3)$$

$$(\widehat{\alpha_{wr}}, \widehat{\beta_{wr}}) = \arg \min_{\alpha_{wr}, \beta wr} \sum_{jt: 0 \leq pr_{jt} \leq h} (W_{jt} - \alpha_{wr} - \beta_{wr} \cdot pr_{jt})^2 \quad (4)$$

The impact of the threat of sanctions is defined as:

$$\widehat{\tau} = \frac{\widehat{\alpha_{yr}} - \widehat{\alpha_{yl}}}{\widehat{\alpha_{wr}} - \widehat{\alpha_{wl}}}$$

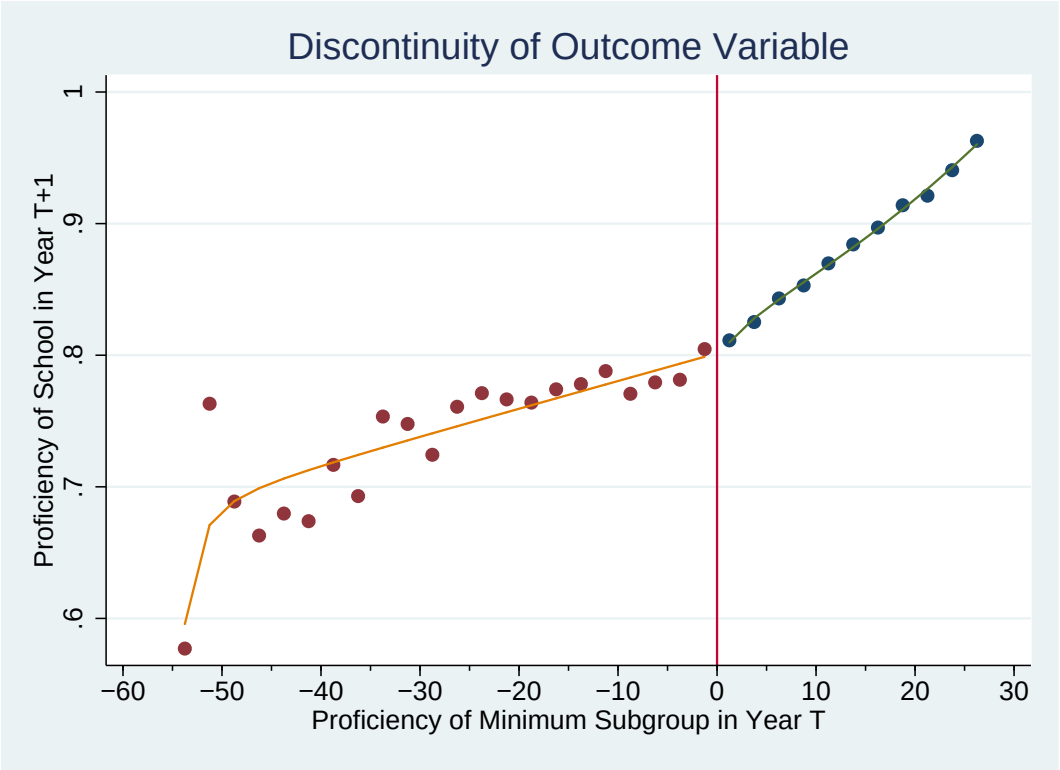
Note that this is equivalent to a local Wald estimator. We expect $\widehat{\tau}$ to be positive and significant if the sanctions had the desired impact.

6 Results

6.1 Aggregate Sanctions

The impact of the treatment is usually visualized by graphing the performance variable (pr_{jt}) against the outcome variable and seeing if there is a discontinuity at the point of treatment. As Figure 3 shows, we do not observe much of a discontinuity in performance, indicating that NCLB sanctions do not improve performance. The results of the RD estimation is presented in Table 4. We do not differentiate between sanctions for this analysis. In an attempt to capture all (if any) impact of sanctions, we tried various outcome measures. The first row results,

Figure 4: School Proficiency Rate in Year t+1



labeled "1-step" results only use the outcome variable and the relative proficiency score of the lowest performing subgroup. The "2-step" procedure first regresses the outcome variable on other school characteristics variables to 'net-out' their impact on outcome variable. That is, in Equations (1) and (2), $Y_{j,t+1}$ is replaced with $\nu_{j,t+1}$ where $\nu_{j,t+1} = Y_{j,t+1} - \gamma Z_{jt}$, where Z_{jt} is a vector of school level characteristics.[5] The "2-step" procedure can be used to eliminate small sample bias and improve precision, but fails to decrease standard errors enough to make any of the sanctions have actual impact on the outcome variables. We note that both procedures are valid methods to perform RD analysis and the identification strategy does not change with the introduction of school characteristics variables.

[5]The complete list of Z_{jt} is: % minority, % students on free or reduced price lunch, average daily attendance, average class-size of 3rd, 4th, and 5th grade students, promotion rates for 3rd and 5th grade students, % of teachers with licensure, average years of teacher experience, % classrooms connected to the internet, ratio of library books to students, and crime rate at the school.

Table 4: Regression Discontinuity Results for Aggregate Threats

Outcome Measure	1-step	2-step
School-wide Proficiency	-0.0608 (0.2782)	-0.0122 (0.1922)
Growth Rate of School-wide Proficiency	-0.0087 (6.4548)	-0.0225 (0.1519)
School-wide Performance	-0.1939 (3.0623)	0.0858 (0.8443)
Growth Rate of School-wide Performance	0.0791 (0.5271)	0.0858 (0.9033)
Worst Fixed Group Proficiency	0.0503 (0.1966)	-0.0277 (9.9907)
Growth Rate of Worst Fixed Group Proficiency	-0.1184 (7.4815)	-0.0558 (0.1771)
Worst Fixed Group Performance	0.3752 (18.2724)	0.1871 (1.3046)
Growth Rate of Worst Fixed Group Performance	-0.4036 (3.2090)	0.3460 (3.8217)
Worst Group Proficiency	0.0283 (0.8853)	-0.0489 (0.4691)
Growth Rate of Worst Group Proficiency	-0.2528 (3.8742)	-0.0889 (0.6500)

6.2 Differentiating Sanctions

It may be that combining all sanctions to evaluate impact may hide effective sanctions. Of the three sanctions that NCLB imposes on schools for failing to make AYP, we can conduct RD analysis on school transfer and supplementary education. There are not enough observations of schools undergoing "corrective actions."

As Table 5 and Table 6 show, neither of the individual sanction threats has any noticeable impact. One note of caution here is that because the sanctions are sequential and additive in nature, the supplementary education threat is on top of the transfer option. That is, schools under threat of being forced to offer tutoring services are already allowing students to transfer to other public schools in the district. Therefore, a 'pure' threat of supplementary education cannot be evaluated in this setup.

7 Specification Testing

[INCOMPLETE]

We also conduct specification tests to make certain that the RD analysis is valid.

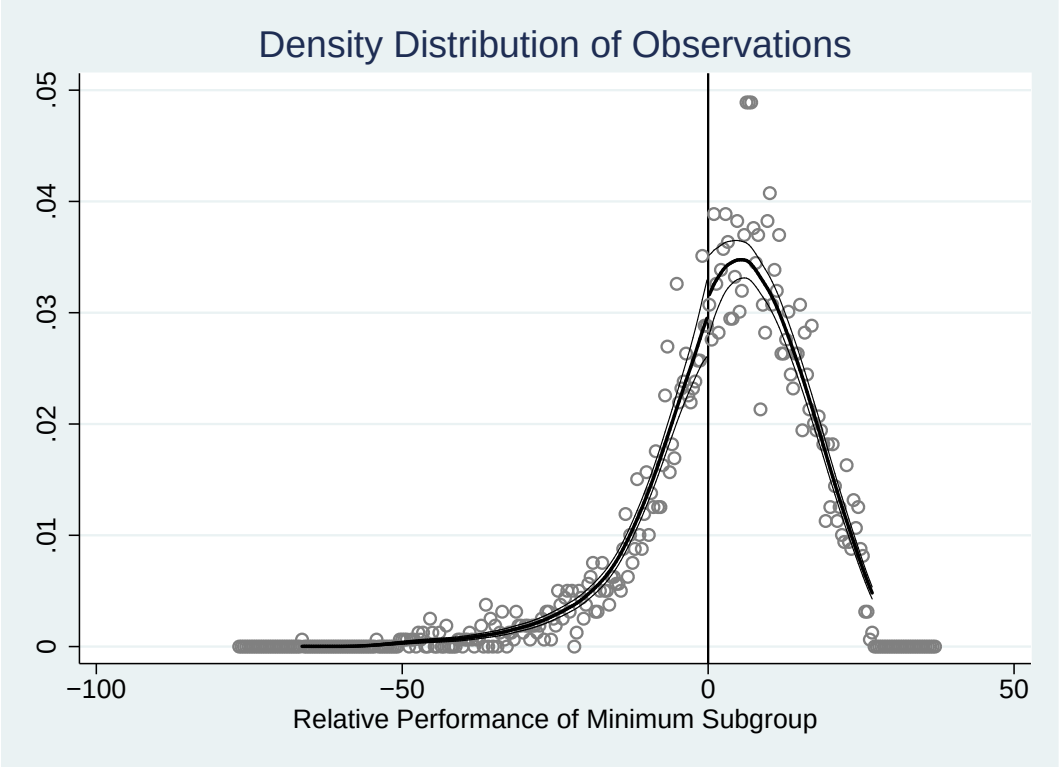
Table 5: Regression Discontinuity Results for Transfer Threat

Outcome Measure	1-step	2-step
School-wide Proficiency	-0.0559 (2.7058)	-0.0438 (0.7835)
Growth Rate of School-wide Proficiency	-0.02826 (0.6985)	-0.0712 (1.2801)
School-wide Performance	-0.1860 (5.6349)	0.0756 (2.1440)
Growth Rate of School-wide Performance	0.0606 (10.2693)	0.0756 (1.3122)
Worst Fixed Group Proficiency	0.2523 (3.6678)	0.0727 (6.4639)
Growth Rate of Worst Fixed Group Proficiency	0.0547 (7.7272)	0.1025 (7.8977)
Worst Fixed Group Performance	0.7571 (22.4387)	0.5836 (9.1889)
Growth Rate of Worst Fixed Group Performance	0.8414 (142.4895)	-0.2268 (52.3917)
Worst Group Proficiency	0.0862 (4.0981)	-0.0541 (6.2462)
Growth Rate of Worst Group Proficiency	-0.2903 (8.7934)	-0.1096 (8.3457)

Table 6: Regression Discontinuity Results for Tutoring Threat

Outcome Measure	1-step	2-step
School-wide Proficiency	-0.2900 (0.2491)	-0.0675 (3.0414)
Growth Rate of School-wide Proficiency	-0.0333 (10.9311)	-0.1134 (5.1967)
School-wide Performance	-1.0672 (31.3299)	-0.0090 (4.1245)
Growth Rate of School-wide Performance	0.1138 (2.8872)	-0.0090 (2.8370)
Worst Fixed Group Proficiency	-0.3159 (9.5819)	-0.3548 (43.3852)
Growth Rate of Worst Fixed Group Proficiency	-0.7255 (12.3136)	-0.5079 (22.5866)
Worst Fixed Group Performance	-0.4169 (5.6456)	-0.4424 (5.3633)
Growth Rate of Worst Fixed Group Performance	-0.7591 (139.9756)	-0.9704 (7.0452)
Worst Group Proficiency	-0.1618 (9.9207)	-0.1670 (134.6462)
Growth Rate of Worst Group Proficiency	-0.5948 (47.7049)	-0.3282 (15.8101)

Figure 5: Continuity of Density



8 Conclusion

[INCOMPLETE]

References

[1] Ahn, T. (2009) "The Missing Link: Estimating the Impact of Incentives on Effort and Effort on Production Using Teacher Accountability Legislation" Working Paper.

[2] Angrist, J. and V. Lavy (2002) "The Effect of High School Matriculation Awards: Evidence from Randomized Trials" NBER Working Paper No. 9389.

[3] Booher-Jennings, J. (2005) "Below the Bubble: 'Education Triage' and the Texas Accountability System." American Educational Research Journal. (42): 231-268.

[4] Bramoulle Y., R. Kranton (2007) "Public Goods in Networks." Journal of Economic Theory (135) : 478-494.

[5] Carnoy M., S. Loeb (2002) "Does External Accountability Affect Student Outcomes? A Cross-State Analysis" Educational Evaluation and Policy Analysis, 24: 305-331.

[6] Clotfelter C., H. Ladd (1996) "Recognizing and Rewarding Success in Public Schools." In: Ladd, H. (Ed.), Holding Schools Accountable: Performance-Based Reform in Education. Brookings Institution, Washington, D.C..

[7] Clotfelter, C., H. Ladd, J. Vigdor (2007) "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." NBER Working Paper 13617.

[8] Clotfelter, C., H. Ladd, J. Vigdor (2007) "Are Teacher Absences Worth Worrying about in the U.S.?" NBER Working Paper 13648.

[9] Cullen, J., R. Reback (2002) "Tinkering Toward Accolade: School Gaming under a Performance Accountability System." University of Michigan Working Paper.

[10] Darling-Hammond L. (2000) "Teacher quality and student achievement: A review of state policy evidence." Education Policy Analysis Archives 8(1)(1 January).

[11] Ehrenberg, R., R. Ehrenberg, D. Rees and E. Ehrenberg (1991) "School District Leave Policies, Teacher Absenteeism, and Student Achievement," Journal of Human Resources (26): 72-105.

[12] Figlio, D., L. Getzler (2002) "Accountability, Ability, and Disability: Gaming the System?" NBER Working Paper 9307.

[13] Figlio, D. (2006) "Testing, Crime, and Punishment." Journal of Public Economics 90: 837-851.

[14] Figlio, D., L. Kenny (2007) "Individual teacher incentives and student performance," Journal of Public Economics, Elsevier, vol. 91(5-6), 901-914.

[15] Goldhaber, D., E. Anthony (2007) Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching. Review of Economics and Statistics, 89(1): 134-150.

[16] Grissmer, D., A. Flanagan (1998) "Exploring Rapid Achievement Gains in North Carolina and Texas." National Education Goals Panel.

[17] Hanushek E., M. Raymond (2005) "Does School Accountability Lead to Student Improvements?" Journal of Policy Analysis and Managment, 24:297-327.

[18] Imbens G., T. Lemieux (2009) "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics, 142: 615-635.

[19] Jacob B., S. Levitt (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Quarterly Journal of Economics, 118(3), 843-77.

[20] Jacob B., (2005) "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." Journal of Public Economics 89: 761-796.

[21] Jacob B., (2007) "Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments." NBER Working Paper 12817.

[22] Ladd, H. (1999) "The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes." Economics of Education Review, Elsevier, vol. 18(1), 1-16.

[23] Milgrom P., C. Shannon (1994) "Monotone Comparative Statics." Econometrica (62) : 157-180.

[24] Muralidharan K., V. Sundararaman (2009) "Teacher Performance Pay: Experimental Evidence from India." NBER Working Paper 15323.

[25] Neal D., (2009) "Designing Incentive Systems for Educators" forthcoming in Performance Incentives: Their Growing Impact on American K-12 Education, edited by Matthew Springer, Brookings.

[26] Neal D., D. Schanzenbach (2009) "Left Behind by Design: Proficiency Counts and Test-Based Accountability." Review of Economics and Statistics, Forthcoming.

[27] Rivkin S., Hanushek, E., and J. Kain (2005) "Teachers, Schools, and Academic Achievement." Econometrica (73) : 417-458.

[28] Rockoff, J. (2004) "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." American Economic Review 94(2): 247-52.

[29] Rothstein J. (2008) "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." Quarterly Journal of Economics, Forthcoming.

[30] Sanders, W. and J. Rivers (1996) "Cumulative and residual effects of teachers on future student academic achievement." Research Progress Report. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

[31] Vigdor, J. (2008) "Teacher Salary Bonuses in North Carolina." National Center for Performance Incentives Working Paper 2008-03.

[32] White K., J. Rosenbaum (2007) "Inside the Blackbox of Accountability: How High-Stakes Accountability Alters School Culture and the Classification and Treatment of Students and Teachers." Forthcoming in No Child Left Behind and the Reduction of the Achievement

Gap: Sociological Perspectives on Federal Education Policy. Eds A. Sadvonik, J. O'Day, G. Bornstedt, and K. Borman. Routledge.