Change for the Better? Education Policy and the Restructuring of American Schools

Travis M. Johnston

August 27, 2012

Abstract

Since its passage, No Child Left Behind (NCLB) has generated contentious debate among school reformers, academics, and politicians alike. Despite this attention, much of the legislation’s impact on schools is still unclear, especially with respect to the restructuring process. Under NCLB, schools that continually fail to meet adequate yearly progress (AYP) are placed in restructuring, a process that often includes a radical transformation of the school’s existing administration, personnel, and educational practices. To the legislation’s supporters, the restructuring process is a necessary step for turning around a perpetually failing school. For NCLB’s detractors, this draconian policy translates to significant upheaval in the classroom and uncertainty in the future. Existing studies of this process have largely failed to produce a reliable estimate due, in large part, to the inherent difficulties in drawing causal inferences from a non-random selection of schools. To this end, I utilize a regression discontinuity design that exploits the sharp AYP cutpoint as a means for comparing schools that just pass the threshold, and are thus safe, to those that do not meet AYP requirements and subsequently receive a restructuring plan. By restricting the analysis to only those schools who narrowly avoid or fall into restructuring, this design is better suited for studying the restructuring process’ effects on subsequent school performance. From a narrow policy view, this study speaks to whether the restructuring process contributes to school turnarounds or merely serves as a further hindrance. Despite finding no statistically significant effect, the study’s null results are substantively relevant to this question of the legislation’s practical impact.
Passed in 2001, the Bush administration’s signature education bill, No Child Left Behind (NCLB), typified the spirit of new federalism. The legislation gave tremendous latitude to states when crafting their educational programs and curriculum, all the while holding tightly to the purse strings. For states that followed the new law’s guidelines, the federal government continued to support public education. Although left to their own devices to develop standards and assessments, the states ultimately operated under the same general philosophy: to achieve and grow academically, schools must make regular testing and evaluation, of both students and schools, a priority.

As part of this new era of accountability, NCLB established systems to track and adequately respond to failing programs. For those perpetually struggling schools, radical change would not only be expected, but required. To this end, schools are now classified according to whether or not they meet adequate yearly progress (AYP), a standardized minimal threshold of academic proficiency. Schools that fail to meet AYP standards two years in a row are placed into a category known as Program Improvement (PI). Once in PI, the school must meet AYP for two consecutive years in order to return to normal status. Schools that do not meet these standards, however, continue to move through the PI system for each additional year in which they fail to meet AYP. Every step in this process brings with it additional requirements and monitoring by larger state and district agencies. Once a school has reached the fourth year of program improvement (PI-4), five years without meeting AYP, they

\footnote{Depending on the school’s particular demographics and levels, each school has many different AYP criteria they must meet. For now, I am only referring to the two major school-wide benchmarks for math and English Language Arts.}
are then classified as a *restructuring school*. It is during restructuring when NCLB’s regulatory teeth are most apparent.

To the legislation’s supporters, the restructuring process is a necessary step for turning around a perpetually failing school. After under-performing for more than five years, advocates claim, schools find it very difficult to make significant gains through minor tweaks here and there. While painful, this band-aid approach offers greater opportunity for dramatic growth in the future, even if the school’s short-term prospects are not the best. Research on restructuring in high schools, finds that graduation rates increase as a result of new accountability measures (Darling-Hammond 2006).

NCLB’s detractors, by contrast, argue that this Draconian policy translates into significant upheaval in the classroom and uncertainty for staff, parents, and students. In the end, restructuring “usually means that little has changed” (Mead 2007). Planning for restructuring, PI-4, let alone implementation of the plan, can throw an already struggling school into complete disarray. Knowing that everyone may lose their job the following year can be, at the very least, quite destructive to a school’s sense of trust and community (Daly 2009). At worst, planning has the potential to instigate a mass departure of the school’s highly sought after teachers and staff. When the school reaches the implementation phase, PI-5, no certainty exists from one day to the next. Moreover, despite having developed a plan the year before, it is rarely implemented in a timely and decisive fashion due to budget and resource constraints.²

² For instance, if the plan calls for a state or charter school takeover, there needs to be resources and partners in place before this can happen. Alternatively, if the plan requires staff firings, as
The tangible effects of restructuring, although of considerable interest in recent years, are still poorly understood. Existing studies of this process have largely failed to produce a reliable estimate due, in part, to the inherent difficulties in drawing causal inferences from this non-random selection of schools. Schools who fall into restructuring are very different from the average school, both academically and demographically (Kim and Sunderman 2005). Consequently, any analysis of restructuring is bound to suffer from the lack of a good comparison group. In order to generate a more accurate treatment effect, we require a reasonable set of control units in which to contrast. While restructuring is by no means assigned at random, it is possible to leverage the selection process in a quasi-experimental sense to arrive at more internally valid comparison.

Simple comparisons between restructured schools and those not restructured are ultimately hampered by the fact that the two groups are vastly different. But what about those schools who were similarly in their third year of program improvement, and thus in jeopardy of falling into restructuring? In addition to being academically similar to those who fell into restructuring (treatment), because restructuring is determined by school-wide AYP rates, we can measure how close each unit came to either receiving or avoiding treatment. To this end, I utilize a regression discontinuity design that exploits the sharp AYP cutpoint as a means for comparing schools that just pass the threshold (control), and are thus safe, to those that do not meet AYP requirements and subsequently fall into restructuring. By restricting the analysis to only those schools who narrowly avoid or fall into treatment, this design is better...
suited for studying the restructuring process’ effects on future school performance.

In the following section, I briefly discuss how restructuring works under NCLB, as well as some of the existing work on restructuring. I then develop a set of hypotheses on the effects of both the planning phase (PI-4) and the implementation process (PI-5). Before getting to the empirical results, I briefly address the data and method used to test these claims. To assess whether these radical measures produce any measurable impact on the aggregate, I examine the effects of planning and implementation on short-term school performance. Not surprisingly, I find that, for those right around the cutpoint, restructuring seems to exert no measurable influence over school performance in the subsequent year. I then perform a similar analysis on whether implementation has any meaningful effect on long-term growth, again finding no clear results in either direction. I conclude with a discussion of the paper’s larger findings and limitations.

**NCLB and the Effects of Restructuring**

In his first State of the Union in February 2001, President George W. Bush unambiguously declared education reform to be his “top priority.” Indeed, education accounted for over 10% of the entire speech in 2001. Rather than simply calling for more education spending, the president pushed aggressively on the need for a reevaluation of how the education operates at both the federal and local level. Eschewing the traditional approaches of his predecessors, Bush asserted,

---

3 The percent of State of the Union spent on education topics was recovered using the “Trend Analysis” application at policyagendas.org.
When it comes to our schools, dollars alone do not always make the difference. Funding is important, and so is reform. So we must tie funding to higher standards and accountability for results. I believe in local control of schools. We should not, and we will not, run public schools from Washington, D.C. Yet when the federal government spends tax dollars, we must insist on results. Children should be tested on basic reading and math skills every year between grades three and eight. Measuring is the only way to know whether all our children are learning. And I want to know, because I refuse to leave any child behind in America.

The resulting bipartisan bill No Child Left Behind set out to do exactly that, passing the following Summer with huge majorities in Congress.

The Restructuring Process

In order for the fiscal federalism experiment to work, Washington needed to maintain a degree of control over the states via monitoring and sanctions. The Program Improvement designation, and the restructuring process are chief among these procedures. By requiring states to regularly test their students, officials can readily assess which schools are failing year after year to meet the minimum levels of adequate yearly progress (AYP). For those schools that are five years out of compliance, PI-4, the state requires the development of a restructuring plan designed to drastically turnaround the school.

Prior to the restructuring phase, the requirements of NCLB are more of a nuisance to school officials dealing with the day-to-day concerns. In PI-4, however,

---

4 NCLB passed the House (384-45) and Senate (91-8), but was not enacted into law until January 2002.
5 Kim and Sunderman (2005) argue that the reliance on AYP heavily disadvantages schools in low-income neighborhoods with diverse populations. By requiring schools to meet AYP for all subgroups, NCLB essentially forces these diverse schools to comply to more AYP criteria, any one of which is sufficient for designating an overall failure to meet AYP.
the looming sanctions are far more tangible as the school, in collaboration with state agencies, begins to develop a “plan” for restructuring. Rather than tweaking existing programs, the alternative governance plan is expected to completely transform the existing school personnel, curriculum, and managing practices. The plan involves a host of radical changes along one of several paths (See Table 1). If the school fails to meet AYP in the subsequent year, the restructuring plan is then implemented (PI-5). Once all AYP requirements have been met for two consecutive years, the school moves out of the program improvement designation.

<table>
<thead>
<tr>
<th>Table 1: Different Restructuring Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reopen as Public Charter School</td>
</tr>
<tr>
<td>Replace Teachers, Staff, and Principal</td>
</tr>
<tr>
<td>Contract with Outside Agency</td>
</tr>
<tr>
<td>State Takeover of School Operation</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

Measuring Restructuring’s Effects

While NCLB’s goal to achieve 100% proficiency by 2014 is laudable, many researchers and practitioners remain skeptical of the legislation’s core approaches. In an evaluation of the law’s fundamental logic, Forte (2009) identifies three relatively unfounded assumptions that NCLB makes. Most importantly, Forte challenges the very idea whether the restructuring process’ “consequences, sanctions, resources, and supports” even produce increased student achievement (2009, 77).

---

6 By far the most common option is “Other,” (CEP 2006). This choice usually includes a combination of different approaches, such as personnel change and private agency partnerships.
In an early examination of the restructuring process in California, the Center on Education Policy finds that considerable time and deliberation goes into planning restructuring (CEP 2006). Yet, despite this attention, school performance post-restructuring has been quite mixed across the state. Similarly, in a recent, macro-level analysis of NCLB’s effects, Dee and Jacob (2011) conduct a longitudinal study of NCLB’s aggregate effects on state achievement. Rather than examining schools or students, they study state-wide performance before and after enactment of NCLB. Dee and Jacob pay particular attention to how the bill’s passage affects those states that had little to no accountability measures in place before NCLB, compared to those states with existing systems. In the end, they find that NCLB produced large gains in math achievement among fourth graders, but not in reading. Once again, NCLB’s effects proved mixed at best.

Adding to this trend of mixed results, Darling-Hammond finds (2006) in a study of restructured high schools that, while beneficial to some metrics such as graduation rates, restructuring can also produce adverse incentives. For one thing, by focusing heavily on the AYP bottom line, the regulations end up discouraging highly qualified teachers from seeking positions at schools with already struggling programs. As a result, this unintended consequence often produces a self-reinforcing cycle, thus making the prospect of post-restructuring success highly unlikely.

In addition to turning away new teachers, Daly (2009) argues that an unwillingness to adapt sets in among the staffs at these struggling schools. Rather than responding to the threat of restructuring by changing the curriculum or management, Daly finds that teachers at PI schools often lack the sense of trust and empowerment
necessary for turning around a failing school. This analysis provides yet another potential mechanism for understanding why restructuring, or even the threat of it, rarely produces any measurable improvement in test scores.

Despite considerable research into the effects of No Child Left Behind, no study to date develops a general estimate of the effects of restructuring on school performance. Without restructuring, the accountability measures implemented by NCLB are meaningless. For NCLB to have an effect, the threat of restructuring should not only motivate struggling programs, but it should also exert an independent effect on those schools in which a plan is implemented. Anecdotal evidence on restructured schools and aggregate studies of state-wide performance are both valuable contributions to understanding NCLB’s impact, but neither can answer the question: does restructuring matter? To understand whether these accountability sanctions increase achievement, either directly or indirectly, we must evaluate their effects on a very specific subset of schools - those units under program improvement who face restructuring in the immediate future.

**Hypotheses**

To assess whether restructuring itself has any direct or indirect effects upon school performance, I plan to analyze only those schools (PI-3) in actual jeopardy of being restructured. For many schools, the mere presence of No Child Left Behind has little import on day-to-day operations. Struggling schools, particularly those in the fourth year of program improvement (PI-4), face the impending threat of a restructuring plan, and thus are most likely to respond, if possible.
**H1:** Threat of restructuring on achievement.  
*a)* If the fear of an impending restructure has an effect on a school’s subsequent performance, and the school has the ability to respond to such accountability measures, then we should expect to see schools in PI-4 (treatment) improve in the next year in an effort to stave off implementation.  
*b)* Alternatively, if restructuring only adds to the school’s anxiety and academic disarray, then those schools who receive treatment should perform worse than those comparable schools who just avoided a restructuring plan (control).

In addition to these indirect effects associated with the threat of restructuring, NCLB also attempts to produce direct effects on school achievement by forcing PI-5 schools to implement the restructuring plan. While unlikely to make a positive effect in the immediate future, NCLB advocates insist that these radical measures are necessary for turning around a perpetually dismal school.

**H2:** Short-term implementation effects. Given increased uncertainty and radical changes to existing structures and personnel, the implementation (treatment) is likely to have no consistent effect on subsequent school performance. If anything, this general upheaval would produce diminished results in the immediate future.

**H3:** Implementation effects on future performance. If implementation of a restructuring plan (treatment) takes time, then school-wide measures of achievement should increase after a considerable period of time.

**Method**

Although well-researched, the question of whether restructuring matters or not is ultimately plagued by problems owing to the lack of a good comparison group. To get around this, most researchers examine how a few restructuring schools change over time, often using numerous qualitative and quantitative measures of performance.
While certainly helpful, this over time analysis is insufficient at producing a causal estimate as many other factors covary with time. How can we be sure that it is the restructuring process itself that leads to the results? To reconcile this problem, the treated units require a good control group to compare against. Unfortunately, there is no way that conditional independence holds for most reasonable control groups. The treated units, either in planning or implementation, are already a very narrow, highly unique set of schools, not to mention their similar cultural and demographic features (Kim and Sunderman 2005).

Despite these challenges, there does exist a set of units well-suited for comparing against the treated observations. Given the structured nature of treatment assignment, the only way a school can be classified as restructuring is by gradually reaching PI year four or five, for cases of implementation. Consequently, we can leverage this process to design a good counterfactual for the treated units - those schools who were similarly in jeopardy of falling into restructuring the previous year. Moreover, because treatment is assigned discontinuously around a known cutpoint, AYP, we can further restrict the sample of control and treated observations to those units right around the cutpoint. Before elaborating more on this approach, I now turn to the data used in the analysis.

**Data**

For this analysis of restructuring’s effects, I used data exclusively from the California Department of Education. The Department of Education has testing and program
improvement data going back to 2002. Given the large number of schools ($N \approx 10,000$) and, unfortunately, history of educational woes, California has a sizable population of schools in program improvement ($N \approx 2,200$ in 2007). For the scope of this study, I will only be examining data from the 2006-2007 school year.

For an overall view of how California schools performed on the 2007 Math and English Language Arts (ELA) exams, see Figure 1. The distributions appear relatively normal, and note that the AYP cutpoints are clearly well below their respective medians. These percent proficiency variables are especially important not only at summarizing the population, but also because the test data ultimately serves as the variable forcing units into treatment or not. For the study’s outcome data, I employed several measures of school performance, including the AYP test scores themselves, but instead settled on the Academic Performance Index (API) score. In addition to its easy interpretation, the API score has a number of good properties for a study of this nature. As a product of test data across the school’s many targets, it provides a nice summary measure of school performance. Moreover, API scores are substantively meaningful to involved parts. As a highly publicized and well-regarded measure of school success, API scores are consciously considered by staff, policymakers, and parents when making decisions about a specific school.

---

7 Focusing on one state is preferable to pooling across several units since states do not use a) the same standards for testing, b) comparable test instruments, or even c) the same AYP benchmarks. Though a federal law, NCLB’s explicit desire to leave planning to local agencies makes a multi-state analysis not only tricky, but also a bit unclear in terms of interpreting results.

8 While I could pool across years, this would prove quite tedious as the same unit may be used on multiple sides of the analysis. In other words, because so many schools eventually fall into treatment, yesterday’s control is often today’s treated unit.

9 The forcing variable, $x$, is the technical term for the covariate that produces the treatment assignment at the cutpoint, $x_0$. See “Identification” below.

10 Scores range from 200 to 1000, with an overall target around 800.
Overall performance levels on Math and ELA tests. The solid red lines indicate the AYP cutpoints for elementary and middle schools. The dashed line indicates the respective high school cutpoints.

**Identification**

With one in five schools in some level of program improvement, California seems well-suited for an evaluation of restructuring. That said, this sample still requires considerable trimming. Because it would be unfair to compare treated units to
schools from earlier stages of program improvement, the following analyses only look at those schools in jeopardy of restructuring (PI-3 = 471) or those that have already developed a restructuring plan but have yet to implement it (PI-4 = 332). Using school-level test data from Spring 2007, schools are then divided into treatment or control depending upon their performance on the state-wide exams. Yet, instead of increasing smoothly across the distribution of proficiency levels, $x$, the designation of restructuring $(0−1)$ changes in a sharp, discontinuous fashion at the AYP cutpoint, $x_0$. This feature of treatment assignment is immensely important for a regression discontinuity (RD) design.

By considering only those units within a narrow “bandwidth” or window around this cutpoint, regression discontinuities provide much firmer grounds for measuring the existence of a local average treatment effect (LATE). If the sample is restricted to a small group of schools just above and below the AYP target, then treatment assignment at the cutpoint can be seen as almost at random.\textsuperscript{11} Studying the average treatment effect of restructuring by comparing all restructured schools to non-restructured schools would be highly problematic. But a local average is much more reasonable when we consider how close either set of units came to receiving or not

\textsuperscript{11} This is of course a highly contested claim, and indeed a central hurdle when making the case that the RD holds. In the world of standardized tests and school reform, we know for a fact that some schools cheat in an effort to make AYP. This is certainly a major concern, for both this paper and education policy more broadly. That said, for the present analysis, this concern is less of a problem if cheating occurs at the classroom level, where the individual teacher performs the fraudulent act. It is a much larger issue if selection into treatment happens in a highly strategic means by which school officials calculate how many students they need to just surpass the school-wide threshold. By and large, incidents of cheating seem to be the former, but we cannot fully rule out the possibility. To that end, the densities of the bins just above and below the cutpoints in Figures 2 and 6 do not suggest any evidence that there is sorting (cheating) going on around the AYP threshold.
receiving treatment.

More formally, if an individual school $i$ fails to meet all AYP targets, $x_0$, in its **fifth consecutive year** (PI-3), then it must develop a restructuring plan the following year, $D_R$.

Assessing Restructuring

$$D_R = \begin{cases} 
1 & \text{if } x_i < x_0 \\
0 & \text{if } x_i \geq x_0 
\end{cases}$$

Similarly, if we want to examine the effects associated with the actual implementation of the plan, then the same logic holds for the PI-4 schools already in restructuring. Schools that fail a **sixth time** must then implement the plan, $D_I$.

Assessing Implementation

$$D_I = \begin{cases} 
1 & \text{if } x_i < x_0 \\
0 & \text{if } x_i \geq x_0 
\end{cases}$$

Unfortunately, identification is not this cut and dry. Traditionally, treatment assignment for a regression discontinuity operates via a single forcing variable.\textsuperscript{12}

To meet adequate yearly progress, NCLB sometimes imposes upwards of twenty conditions that school $a$ must surpass. Schools must not only meet AYP benchmarks on each subject matter test (ELA and Math), but they must also be met by every relevant subgroup at the school (e.g. low-income, racial minority, special education, etc.). NCLB also imposes a participation rate for each test and subgroup to ensure

\[\text{12 For an overview, see Lee and Lemieux (2010).}\]
that the school is not strategically testing only those students expected to pass.\textsuperscript{13} In addition to the testing and participation targets, schools have to also meet a minimum API growth threshold and, for high schools, a graduation rate.

An attempt to analyze all these separate forcing variables would prove intractable, particularly since many of these criteria only matter in a few cases. Instead, I have confined my analysis to the numerically largest cutpoint of interest: ELA at the school-wide level.\textsuperscript{13} When looking at the raw data in Figure 2, it is clear that the ELA target was a much more difficult hurdle than the math target. Moreover, among those schools that failed to meet either of the school-wide thresholds, the distance from the ELA target was typically greater than the math distance.

With this very muddled set of multiple forcing variables, an analysis centered around the ELA cutpoint is also the most methodologically sound choice. If we are to select a metric for assessing how close a unit came from crossing the threshold, it seems reasonable that the measure with the greater distance should generate a more conservative estimate. Hence, in the following analysis, I present discontinuity results for those schools that failed both the Math and ELA targets, or just latter. By restricting the analysis to these observations, I can use the same cutpoint for all units, thus allowing for a more intuitive interpretation of how much a given unit was above or below the treatment threshold.

\textsuperscript{13} Generally speaking, most schools pass the participation thresholds easily enough. Usually this simply entails an aggressive grass roots campaign to guarantee that students are not absent during testing periods.

\textsuperscript{14} By and large, the findings of the paper (null results) hold for the math cutpoint results as well. For the sake simplicity, the specific subgroups that make up a school’s overall rate are set aside as well.
The two histograms above show how far the units were from their relevant cutpoints. The distance metric was normalized to reflect the school-wide percent above or below the AYP target for that particular subject. The unit of analysis is PI-3 schools in 2007. For a similar plot for PI-4 schools, see Figure 6.

Results

To measure the disruptive, short-term impact of falling into restructuring ($H1$), the first analysis seen in Figure 3 focuses on whether the development of a restructuring
plan in PI-4 has any indirect effect on the year’s test results. If a school hopes to avoid restructuring implementation, then we should see a marked improvement between those units just below who received treatment, $D_R = 1$, and those right above the AYP cutpoint, $D_R = 0$. Alternatively, if the restructuring phase causes greater disarray, then we should expect treated units to perform worse than those schools right above the AYP threshold.

Figure 3: Restructuring Effects on Short-term Performance - T+1
As Figure 3 clearly shows, there is absolutely no deviation apparent in the units right around the cutpoint. Rather than increasing or decreasing at the threshold, the data are remarkably linear throughout. Indeed, the two loess smoothers on either side of $x_e$ not only line up, but they also map perfectly onto the full sample smoother.\footnote{Loess is a locally weight smoother that non-parametrically fits the data using a nearest neighbor weighting approach. I tried several specifications, all of which support a null finding. Included plots use \texttt{loess} function in R’s stats package.} While by no means conclusive, the smoothness across the threshold strongly suggests that planning for restructuring has no effect upon subsequent school performance. Moreover, the absence of any jump or perturbation, hence a treatment effect, is further corroborated by the regression results in Table 2. Restructuring, while highly significant in a bivariate model, loses all significance when the lagged dependent variable is included. In the third model, restructuring, $d$, is again insignificant when running a linear model that includes the forcing variable, $x$, distance from the ELA cutpoint.\footnote{Polynomial regressions consisting of cubic and quadratic models were also considered, but of course these models seem unnecessary given the apparent lack of any discontinuity in the data.}

The second and third analyses explore the effects of implementation itself, first on short-term prospects, and then in analysis of the outcome variable several years out. Like the restructuring PI-4 analysis, Figure 4 plots the outcome variable, API 2008, against distance from the ELA cutpoint in 2007, the year treatment is assigned. As with the previous plot, this figure exhibits no discontinuity whatsoever around the AYP threshold, $x_e$. In other words, like the PI-4 schools, the data provide no evidence that implementation affects school performance, neither visually in the graphical plot nor numerically in the regression results (see Table 3).
To assess the effects of implementation on future performance, I conduct an analysis similar to that of the short-term cases, but instead evaluate outcome data several years after implementation. While set up in an identical fashion, the delay between treatment and outcome, API scores in 2011, presents a challenge in that many of the control cases, $D_t = 0$, are not controls during the entire intervening period. Because a school must meet AYP for two consecutive years to exit program.
improvement, avoiding treatment in 2007 does not necessarily imply that treatment was avoided for good.\textsuperscript{17} Instead of simply fitting loess smoothers on both the treated and control groups in their entirety, Figure 5 uses only those control units (highlighted in blue) that remained control throughout. In either case, we once again find no evidence of a discontinuity around the cutpoint. Yet, unlike the previous two plots, \textsuperscript{17} More formally, we are concerned here about units in which $D_T = 0$ in $T+1 \rightarrow D_T = 1$ in $T+3$.}
this analysis uses a considerably smaller sample, and is thus limited by the lower power.

**Discussion**

This study’s results, while lacking statistical significance, hopefully provides some new evidence in the debate on NCLB’s restructuring process. Despite failing to find a significant result, the substantive implications are not necessarily wrong in every case. Hence, a finding of no effect may be true, particularly in the first two plots where the statistical power is higher. Traditionally, a null result does not necessarily indicate a confirmation of no relationship. That said, the regression discontinuity plots appear so strikingly smooth and continuous, that it would be very difficult to claim that restructuring has a consistent effect on school performance. Indeed, if there is a true effect of restructuring, it is likely to be conditional on other contextual factors.

Before addressing how this analysis could be improved and extended, a word on the many limitations to this study. First, the power of the tests, while good in some cases, is ultimately too low to produce a stable estimate. Given the huge variance within the population, the sample size is certainly a challenge when trying to find a local average treatment effect. Moreover, because the $N$ was so low, I resorted to using local regression on the units themselves. In an ideal setting, these individual cases should have been placed into bins designated by their distance from the cutpoint, and then analyzed from there. Second, the finding of null results, while
interesting to those school reformers that believe restructuring has little impact, is again challenged by the sensitivity of the outcome data. To get into the study’s sample a school had to have been struggling for years. As a result, the API scores and future ELA test results are remarkably stable from one year to the next, thus making a significant finding one way or the other difficult. Put another way, if failing is path-dependent, as few would dispute, then the ability to register a modest gain or loss is all that much harder. Consequently, it is possible that a no relationship does exist, but it is also likely that restructuring exerts a small effect not easily captured by these large aggregate measures.

In an attempt to solve some of these issues of power and sample size, this study has a few logical next steps that I hope to explore soon. First, by pooling units over time, I may be able to generate a large enough sample to feel more comfortable with both estimates of main effects and yet to be explored subgroups. Because of the dynamic manner in which treatment is assigned, units move in and out of restructuring over time. This presents a challenge for inference, but should be considered more fully, especially when examining the long-term effects of implementation. Second, with a larger sample, I can then investigate whether restructuring produces heterogeneous effects conditional on a school’s specific demographics, history, or even the type of restructuring implemented. Subgroups, such as low income or highly racially diverse schools, may share similar patterns after restructuring. Alternatively, how long a school has been struggling, or whether the school cannot meet numerous AYP thresholds may also produce similar trends when analyzed separately. Lastly, the nature of the treatment, or restructuring plan itself, may also hold some promise
for understanding implementation effects. As Table 1 shows, schools have several options to choose from, and it is quite likely that a specific combination generates conditional effects, but is swamped by the full sample analysis.

In the end, it seems unlikely that any of the hypotheses are necessarily right or wrong. Given the study’s statistical limitations and unexplored subgroups, we cannot confidently conclude that restructuring has no effect. We should, however, be leery of claims that restructuring produces stable and consistent benefits or drawbacks. Rather than registering any significant jump or fall at the threshold, schools who just receive treatment looked nearly identical to those comparable schools who narrowly avoided NCLB’s much maligned educational intervention. While perhaps impossible to conclude whether restructuring affects subsequent school performance, it is unlikely that the policy and similar accountability measures are going anywhere for the time being (Mintrop and Sunderman 2009).

Works Cited


Appendix

Figure 6: Distance From AYP Cutpoints

The two histograms above show how far the units were from their relevant cutpoints. The distance metric was a normalized to reflect the school-wide percent above or below the AYP target for that particular subject. The unit of analysis is PI-4 schools in 2007.
Table 2: Regression Results for $D_R$ on 2008 API

<table>
<thead>
<tr>
<th></th>
<th>Bivate - Full</th>
<th>Lagged DV - Full</th>
<th>Linear - RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>727.18 ***</td>
<td>92.94 ***</td>
<td>665.54 ***</td>
</tr>
<tr>
<td>$d$</td>
<td>-53.42 ***</td>
<td>-1.63 (3.57)</td>
<td>10.39 (9.71)</td>
</tr>
<tr>
<td>$API07$</td>
<td>0.89 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x$</td>
<td></td>
<td>5.84 *** (0.46)</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>426</td>
<td>425</td>
<td>249</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.10</td>
<td>0.83</td>
<td>0.64</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.10</td>
<td>0.83</td>
<td>0.63</td>
</tr>
<tr>
<td>Resid. sd</td>
<td>58.94</td>
<td>26.02</td>
<td>43.01</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
† significant at $p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

Table 3: Regression Results for $D_I$ on 2008 API

<table>
<thead>
<tr>
<th></th>
<th>Bivate - Full</th>
<th>Lagged DV - Full</th>
<th>Linear - RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>712.36 ***</td>
<td>83.40 ***</td>
<td>673.57 ***</td>
</tr>
<tr>
<td>$d$</td>
<td>-38.24 ***</td>
<td>3.55</td>
<td>-3.59 (10.50)</td>
</tr>
<tr>
<td>$API07$</td>
<td>0.90 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x$</td>
<td></td>
<td>4.17 *** (0.64)</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>320</td>
<td>320</td>
<td>178</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.12</td>
<td>0.74</td>
<td>0.50</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.12</td>
<td>0.74</td>
<td>0.49</td>
</tr>
<tr>
<td>Resid. sd</td>
<td>43.20</td>
<td>23.63</td>
<td>35.21</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
† significant at $p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

Table 4: Regression Results for $D_I$ on 2011 API

<table>
<thead>
<tr>
<th></th>
<th>Bivate - Full</th>
<th>Lagged DV - Full</th>
<th>Linear - RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>722.36 ***</td>
<td>275.16 ***</td>
<td>715.58 ***</td>
</tr>
<tr>
<td>$c$</td>
<td>28.02 **</td>
<td>-0.82 (8.14)</td>
<td>6.13 (9.26)</td>
</tr>
<tr>
<td>$API07$</td>
<td>0.68 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x$</td>
<td></td>
<td>2.70 *** (0.35)</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>307</td>
<td>307</td>
<td>307</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.03</td>
<td>0.36</td>
<td>0.19</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.02</td>
<td>0.36</td>
<td>0.18</td>
</tr>
<tr>
<td>Resid. sd</td>
<td>49.25</td>
<td>40.02</td>
<td>45.14</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
† significant at $p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$