NATIONAL CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

School Turnaround in North Carolina: A Regression Discontinuity Analysis

Jennifer A. Heissel
Helen F. Ladd

WORKING PAPER 156 • March 2016

School Turnaround in North Carolina: A Regression Discontinuity Analysis

Jennifer A. Heissel
Northwestern University

Helen F. Ladd
Duke University

Contents

Acknowledgements

This research was made possible in part by generous support from the National Center for the Analysis of Longitudinal Data in Education Research (CALDER). Research supported by CALDER is funded through Grant R305C120008 to the American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education.

We thank David Figlio and participants at the Association for Education Finance and Policy annual conference for their comments on earlier versions of this paper. We also thank the New Teacher Center for their help with the North Carolina Teacher Working Conditions data.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

School Turnaround in North Carolina: A Regression Discontinuity Analysis
Jennifer A. Heissel & Helen F. Ladd
CALDER Working Paper No. 156
March 2016

Abstract

This paper examines the effect of school turnaround in North Carolina elementary and middle schools. Using a regression discontinuity design, we find that turnaround led to a drop in average school-level math and reading passing rates and an increased concentration of low-income students in treated schools. We use teacher survey data to examine how teacher activities changed. Treated schools brought in new principals and increased the time teachers devoted to professional development. The program also increased administrative burdens and distracted teachers, potentially reducing time available for instruction. Teacher turnover increased after the first full year of implementation. Overall, we find little success for North Carolina's efforts to turn around low-performing schools under its federally funded Race to the Top grant.

1. Introduction

Programs to “turn around” consistently low-performing schools have sprung up in states across the country, bolstered by the federal No Child Left Behind and Race to the Top programs. The schools at the heart of these initiatives face problems ranging from low test scores to student behavior problems, poor school leadership to high staff turnover rates. The persistence of their problems and the fact that such schools typically serve high concentrations of low income and minority students have made turning them around a central part of the federal government’s recent efforts to improve education. A key aspect of the school turnaround strategy is the view that piecemeal reforms related to particular inputs such as teacher qualifications or class sizes will not solve their problems. Instead what is needed, according to this view, are broader whole-school reform efforts that address in a more comprehensive way the range of problems such schools face such as weak leadership, low teacher morale, low expectations for students, and poor school climate. Despite little rigorous research on the potential for the school turnaround approach in recent years, the federal government leveraged its limited funding for education – funding that was temporarily greatly enhanced with post-recession stimulus dollars after 2009 – to induce states to adopt one of four clearly specified school turnaround strategies to improve their lowest performing schools.

This paper contributes to the surprisingly limited body of rigorous research on the school turnaround approach by examining a federally supported program in the state of North Carolina, called “Turning Around the Lowest Achieving Schools” or TALAS. Because the state used a clear cut off to identify the schools to be turned around we can use a regression discontinuity analysis to determine the causal effects of the state’s program. North Carolina is particularly interesting for this study because the state has been surveying all teachers in the state biannually for many years. Information from these surveys makes it possible to investigate not only how the state’s turnaround model affected student outcomes, but also the potential mechanisms through which the program exerted its influence on the schools.

A major purpose of the state's TALAS program was to improve student outcomes, with the specific goal of improving school level student passing rates by 20 percentage points in the turnaround schools (RttT Application, 2010). We find, however, that the turnaround program did not increase average achievement at either the school or the student level. Instead it appears to have reduced overall passing rates in the treated schools. Although we cannot pinpoint the specific causes of this disappointing outcome, we are able to isolate a number of both intended and unintended changes at the school level that could have contributed to it.

1.1 Background and Prior Research

Most turnaround programs seek to improve student achievement in low-performing schools by changing their leadership and culture. The general consensus appears to be that lasting change requires changes in principal and teacher behavior in schools, whether through staff turnover or professional development. Many turnaround programs specifically call for firing the principal. Principals are particularly important to schools, as they make personnel decisions, set policies and practices, and influence school culture. Principals vary in their effectiveness, especially in higher-poverty schools (Branch, Hanushek, & Rivkin, 2012). The effect of changing principals in turnaround schools, however, depends on the relative quality of the replacing principal. Limited experience as a principal is predictive of low school performance (Clark, Martorell, & Rockoff, 2009), and replacing an ineffective, experienced principal with an unknown, inexperienced principal may bring few benefits and could be counterproductive.

Principals can also encourage a distributed model of leadership. Distributing leadership functions across a school results in the school-wide capacity-building and ownership needed to sustain school reforms (Copland, 2003). Turnaround schools may benefit from a combination of transformational and instructional leadership, both of which are viewed as necessary but insufficient for success (Marks & Printy, 2003). Transformational leaders change school culture, emphasize innovation, and support and empower teachers as part of the decision-making process. Shared instructional leadership involves active teamwork between the principal and teachers on curriculum, instruction practices, and student assessments (Marks & Printy, 2003). Autonomy from local control may also help schools improve; in the United Kingdom, schools that narrowly voted to become autonomous schools funded directly by the central government posted large achievement gains, relative to schools that narrowly voted against the change (Clark, 2009).

Principals also influence school quality through their personnel decisions (Branch et al., 2012). It is well known that many teachers tend to avoid schools serving minority and low-income students, and these disparities systematically affect student performance (Boyd, Lankford, & Wyckoff, 2007; Clotfelter, Ladd, & Vigdor, 2007, 2010; Hanushek, Kain, & Rivkin, 2004; Jackson, 2009). But studies also show that even after researchers control statistically for student demographics, teachers' decisions to remain in a school are also strongly influenced by the working conditions in the school, a major determinant of which is the quality of the school's leadership (Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005; Moore Johnson, Kraft, & Papay, 2012).

In addition to principal change, some turnaround efforts also require schools to replace 50% of their teachers. The usefulness of this policy depends on the quality of the replacement teachers. Such a requirement makes little sense for rural areas where there is a limited supply of qualified teachers to replace those who are fired (Cowen, Butler, Fowles, Streams, & Toma, 2012; Sipple & Brent, 2007). Alternatively, there is some evidence that teachers can improve their joint productivity in low-performing schools (Hansen, 2013). Many programs attempt to create these improvements through professional development, but to create improvements the programs must be of high quality. Many studies document that the standard one-shot programs not related to the curriculum do not make teachers more effective (Garet et al., 2008, 2011).

Despite the growth of school turnaround efforts that include these or other components, little research has examined their causal effects on student outcomes. A review by the What Works Clearing House in 2008, for example, found no studies of turnaround programs that met their standards for internal validity (Herman et al., 2008). A more recent review found that fundamental cultural transformations are quite difficult, particularly with a short window of funding (Anrig, 2015). The most careful causal study in the United States to date is a regression discontinuity study of school turnaround programs in California (Dee, 2012). Dee finds that the program significantly improved the test scores of students in low-achieving schools, particularly among schools that replaced the principal and at least 50% of the staff. One limitation of this study is that it was based on a competitive federal School Improvement Grant program, with only about half of the eligible bottom 5% of schools receiving turnaround funding. The concern is that the schools (among the lowest-performing schools) with the best available staff or most supportive districts were the ones to apply for and receive funding. Hence, the positive findings might not apply to the typical low-performing school.

1.2 North Carolina Policy Context

North Carolina has been engaged in school turnaround efforts for almost 10 years. Created in 2006, the District and School Transformation department, or DST, focused efforts on the 66 lowest-performing high schools to increase student achievement. The program expanded to 37 middle schools in 2007. All schools received some support, but these schools received a transformation coach, instructional facilitators to provide instruction and classroom-level support, and a reform or redesign plan (Department of Public Instruction, 2011). The interventions were most intensive in high schools, where they were judged to have modest but significant positive effects on student test scores (Thomson, Brown, Townsend, Henry, & Fortner, 2011). Drawing on that experience, the state successfully competed for federal Race to the Top Funds to turn around the lowest 5 percent of the state's schools. The analysis in the current paper focuses on this recent program – Turning Around the Lowest Achieving Schools, commonly called TALAS – that began in 2011.

Although TALAS also applies to high schools, we limit our analysis to the 85 elementary and middle schools that were subject to this program. High schools did not have the same clean assignment cut point as younger schools, as graduation rates also factored into their assignment. Leaving out high schools also reduced the potential for confounding the effects of TALAS with the more intensive high school intervention from the previous program. However, given the regression discontinuity design that we employ, we can still make causal claims as long as the TALAS cutoff does not exactly overlap with previous cutoffs.

Per federal guidelines, each TALAS school had to implement one of the US Department of Education's four federal models in the schools (Department of Public Instruction, 2014):[1]

Transformation model: Replace the principal; take steps to increase teacher and school leader effectiveness; institute comprehensive instructional reform; increase learning time; create community-oriented schools; provide operational flexibility and sustained support.

Turnaround model: Replace the principal and rehire no more than 50% of the staff; take steps to improve the school as in the transformation model.

Restart model: Convert the school or close and reopen it under new management.

School closure: Close the school and enroll the students who attended that school in other schools in the district that are higher achieving.

By the end of the 2011 school year, all 118 schools in TALAS (including the high schools) had implemented some steps of an intervention model, but many of these had not yet been fully implemented (Whalen, 2011). The majority of schools opted for the transformation model, which required that the principal be replaced but did not require the firing of teachers. That summer, the state introduced an induction and mentoring program for new teachers, as well as three Regional Leadership Academies for principals (Duffrin, 2012). In the 2012 school year, district, school, and instructional coaches provided customized support and professional development to TALAS schools, though turnover in the coaching staff presented problems in the continuity and quality of the training the schools and principals received (Department of Public Instruction, 2013b; Henry, Campbell, Thompson, & Townsend, 2014). Coaches generally served more than one school, with an average of about one day per week spent at a given turnaround school (Henry et al., 2014). The particular strategies employed by the coaches differed by school.[2] In general the leadership coaching strategies employed in turnaround schools did not differ substantially from those used by mentors in non-turnaround schools, though meetings were more frequent (Henry et al., 2014). Required annual progress reports discuss the professional development provided to principals and teachers, with a particular emphasis on school and teacher leadership, as well as principal/teacher recruitment efforts (Department of Public Instruction, 2013b, 2014).[3] Schools continued these strategies in the 2013 and 2014 school years. Our analysis follows schools, students, and teachers through 2014.

2. The North Carolina Data

This paper uses data from K-8 schools in the 2010 through 2014 school years from NCDPI and the North Carolina Education Research Data Center, as well as the 2010, 2012, and 2014 iterations of the North Carolina Teacher Working Conditions Survey. We exclude private, charter, alternative, and special education schools, as they were not eligible for TALAS.

North Carolina started its biannual Teacher Working Conditions survey in 2002. The survey asked questions designed to elicit educators' time use (in ranges of hours per week) and impressions of school climate (on an agree-disagree 4- or 5-point scale). From 2010 to 2014, the individual-level teacher response rate averaged over 90%.[4] We separately analyze the time use and school climate measures. Using the 2010 baseline data, we collapse the school climate data into seven factor composites for teachers' perceptions of their working conditions: leadership, instructional practices, professional development, community relations, student conduct, school facilities and resources, and time use. This method resulted in a Z-score (with an average of zero and a standard deviation of one) for each factor in each school by year. See the Appendix for more details on the survey questions and factor analysis for the school climate data.

For each school in each year, our data include the school-level passing rates for end-of-grade (EOG) tests; student-level test scores and passing rates; and school characteristics such as the principal of record, one-year teacher turnover, percent of teachers with three or fewer years of experience, student behavior, and student demographics.[5] Students are required to complete EOG tests in reading and math in grades 3-8 and in science in grades 5 and 8. We assume that schools that disappear from the NCDPI data closed.

Assignment to treatment was based on a school's 2010 composite score, which is the percent of reading, mathematics, science, and end-of-course test passed out of all such tests taken in a given school. The bottom 5% of schools in each school type (elementary, middle, and high school) were to be placed in the TALAS program, with additional high schools placed in the program based on low graduation rates.

The baseline sample includes 89 treated elementary and middle schools, which account for 5% of the 1,772 North Carolina public elementary and middle schools eligible for TALAS in 2010.[6] Four treatment schools closed in 2012, one closed in 2013, and one closed in 2014. Several control schools closed as well, leaving 83 treatment schools out of 1,753 schools (4.7%) that were open from 2010 through 2014. In the following analysis, we require schools to appear in all years 2010-2014 to be included in the analysis, though including schools before they closed does not change our results.

3. Estimation Strategy

We estimate the effect of the TALAS program by comparing outcomes for schools just below and just above the discontinuity in treatment created by the 2010 composite score assignment rule. Central to our regression discontinuity (RD) design are the clear cut points that determine which schools are treated under TALAS. The cut points for elementary and middle schools are 52.5% and 54.2%, respectively; they differ slightly to ensure that 5% of each school type is included in TALAS. By centering each school's composite score around the applicable cut point and labeling that 0, we can pool them into a single analysis. Figure 1 displays the treatment uptake by the 2010 baseline score by school type and overall.

The main takeaway from Figure 1 is the strong discontinuity in uptake at the cutoff. We note, however, that two schools above the cut point did not comply with their assignment. It is not clear how two elementary schools above the elementary school cutoff received treatment, though we note that their scores are below the middle school cutoff. These schools may have been misclassified as middle schools in the assignment process. Given the ambiguity of the process, we use a "fuzzy" regression discontinuity (Campbell, 1969) as we explain below. The intended treatment population includes those below the cutoff; the intended control population includes those above that point. This simple comparison provides an intent-to-treat estimate; scaling up the estimated difference by dividing by the compliance rate provides a treatment-on-the-treated estimate.

This regression discontinuity (RD) design builds on the observation that whether a school is just above or just below the cut point is essentially random. One potential concern is that schools may manipulate their baseline scores (Lee & Lemieux, 2010) and in effect choose to receive treatment or not. Given that NCDPI determined the cut point after students took the 2010 baseline assessments (Conaty, 2011), such behavior seems highly unlikely. Moreover, as long as schools, even while having some influence, cannot precisely control the assignment variable, variation near the treatment will still be randomized much like a randomized experiment (Lee & Lemieux, 2010).[7] In any case we find no empirical evidence of such manipulation.[8]

One way to confirm that assignment at the cutoff is "as good as random" is to check for discontinuities at the cut point in various baseline characteristics, including the assignment variable. Table 1 displays both the average value of various baseline characteristics above and below the cutoff (Panel A) and the estimated value at the cutoff point (Panel B). This analysis uses the same parametric function we describe in Section 3.2. Panel A shows that schools below the cutoff have lower average composite scores, higher proportions of free and reduced price lunch (FRL) and Black students, lower average daily attendance, more short term suspensions, and higher teacher turnover than schools above the cutoff, patterns that are expected given the well documented relationship between student test scores and various measures of disadvantage. These differences indicate that a simple comparison of schools above and below the cutoff would produce biased estimates of the effects of the policy intervention. When we focus on a comparison of schools at the cutoff point (as in Panel B), however, the differences disappear.

Although the RD approach provides a strong case for causality, it has three potential limitations. First, it identifies treatment effects only at the discontinuity cutoff, which limits generalizability if treatment effects are not constant across the assignment variable. At the cutoff, however, the estimates can be similar to those in randomized experiments (Lee & Lemieux, 2010; Shadish, Galindo, Wong, Steiner, & Cook, 2011). Moreover, generalizability away from the cutoff might not be a concern in the context of school turnaround, as program expansion would occur at the margin. We note, though, that a finding of either a negative or null effect at the cut point would not rule out a more positive effect on the schools well below the cut point.

Second, specifying the correct functional form presents a challenge. Because we cannot know the "true" functional form in our analysis, RD depends on functional form assumptions, whether parametric or nonparametric. We present a variety of specifications for each outcome of interest, using both nonparametric and parametric methods (Lee & Lemieux, 2010).

Third, RD has much less statistical power than a randomized experiment (Goldberger, 1972; Schochet, 2009). Although in theory we should use the smallest bandwidth possible around the cutoff to arrive at the least biased estimates, shrinking the bandwidth simultaneously decreases the power of our analysis. We balance these considerations by estimating models with varying bandwidths. Intuitively, using schools at the very top of the score distribution as a comparison does not tell us much about what would have happened to schools in the bottom 5% of schools. We use +/-16 percentage points as our largest bandwidth in our parametric analysis, as this size includes all but two treated schools, allows us to divide our sample into two-percentage point bins, and balances the distance from the cutoff available for the treated and untreated populations. In some cases we also report results based on bandwidth of +/-10 percentage points bandwidth. We review our methods in more detail below.

3.1 Nonparametric Estimation

Our "nonparametric" estimates are in fact a series of local linear regressions performed at various bandwidths on either side of the cutoff. We use the optimal bandwidths proposed by Imbens and Kalyanaraman (IK, 2011) as our preferred bandwidth. Using Stata's program *rd* (Nichols, 2011), we specify a triangular kernel, which tends to be the most accurate at the frontier (Fan & Gijbels, 1996). The IK bandwidths differ between estimates depending on the relationship between the assignment variable and the outcome variable. We use the full range of data in this analysis (N=1,753 schools).

3.2 Parametric Analysis – School-Level Analysis

We implement a fuzzy RD design with a two-stage parametric model that functions as an instrumental variable analysis (Hahn, Todd, & Van der Klaauw, 2001; Lee & Lemieux, 2010; Van Der Klaauw, 2008). The first-stage model estimates the jump in treatment probability at the cutoff point, with the following general form:

$$(1) \qquad Turnaround_s = \alpha\, I(A_s \leq 0) + f(A_s) + \gamma X_s + \nu_s$$

where $f(A_s)$ is a function of school *s*'s baseline assignment variable and (X_s) represents baseline control variables. The function $f(A_s)$ is allowed to differ on each side of the cutoff. Because the discontinuity essentially functions as random assignment, including baseline covariates is not strictly necessary (Lee & Lemieux, 2010); we include them in practice to reduce sampling variability.[9] The coefficient α represents the percentage point increase in the probability of receiving treatment at the cutoff. We estimate the 2SLS estimate of the effect of this jump in continuity with the following:

$$(2) \qquad Y_s = \pi\, \widehat{Turnaround}_s + g(A_s) + \beta X_s + \varepsilon_s$$

where Y_s is the outcome of interest regressed on the predicted probability of receiving the turnaround treatment, a function of school's assignment variable $g(A_s)$, and the control variables included in Model 1. Under assumptions of monotonicity (that is, no individuals are *less* likely to take up treatment if they are assigned to it) and excludability, this system of equations functions as an instrumental variable estimate and its estimand, π, should be interpreted as a *local average treatment effect* (LATE, Angrist, Imbens, & Rubin, 1996; Angrist & Pischke, 2009; Hahn et al., 2001). In other words, the estimate is only for those whose uptake is affected by the assignment around the cut point.

Because we do not know the "true" relationship between the outcome and the assignment variable, we cannot be certain whether $f(A_s)$ and $g(A_s)$ should be linear, quadratic, cubic, or something else entirely. Lee and Lemieux (2010) suggest a test to find the best-fitting parametric form.[10] The models that follow use the simplest model not rejected by this test; the vast majority have a linear spline on either side of the cutoff.

3.3 Parametric Analysis – Student-Level Analysis

We use longitudinal data for individual students who were in third or, for some of our models, also sixth grade in a school +/-16 percentage points from the cut point in 2010. We limit the population to these grades because they are the most likely to remain in the same school after implementation in

2012. Fourth and fifth graders likely moved to middle school by 2012, while seventh and eighth graders likely moved to high school. The analysis does not restrict the students to schools that remained open 2010-2014 in order to follow students as they move between available public schools.

The first stage predicts the probability of the student's third grade school receiving treatment based on their 2010 composite score. The second stage predicts the outcome of interest. This is the same as asking, given that your 2010 school received treatment, how did you do relative to a student whose 2010 school did not receive treatment? Students who change schools across years continue to be assigned to their baseline school. The analysis could also be considered an intent-to-treat analysis, with the note that the first stage accounts for the small fuzziness of the assignment at the school level.[11] This student-level approach is limited to one cohort of students, but it avoids potential interpretation challenges related to compositional changes in schools, as we follow the students regardless of the school they attend. We follow students whether they are retained or skip a grade, as long as they remain in a public school in North Carolina. Robust standard errors are clustered by the 2010 school.

Additionally, we can examine outcomes based on how far students were from passing in 2010. In the baseline year, North Carolina placed students in four categories based on their test scores: Levels I and II did not pass, and Levels III and IV passed. This subgroup analysis permits us to determine how the turnaround program affected students with different levels of pretreatment academic performance.

We now turn to our results. We first examine whether student outcomes improved. We then use several outcome measures to try to understand the patterns we observe in the student outcome data. In the results below, we label our nonparametric estimates as NP and our parametric estimates as 2SLS.

4. Did student outcomes improve?

A major objective of the TALAS program was to improve student outcomes, with the specific goal of improving school-level composite scores by 20 percentage points. Thus, the first question we ask is whether the program succeeded in raising student achievement or improving other student outcomes.

We answer this question using two approaches. The first and most central approach uses the school as the unit of observation and examines the patterns of composite scores in math and reading passing rates, as well as student behavior through 2014. In the formal part of this school-level analysis, we report results by student demographic subgroups for the years 2012, 2013 and 2014. The second approach uses student-level data for students who were third graders in 2010. We do not include sixth graders because by 2012 some students were able to select into different tests. Some eighth graders took the EOG math and reading tests, while others took the Algebra I or English I EOC.

The patterns for the most straightforward models, which are depicted in Figure 2 for school outcomes in 2014, indicate that the program had a negative effect on test scores in math and reading. The graph displays the 2010 baseline trend (in gray), the 2014 segment that was intended as a control (in solid black), and the 2014 segment that was intended for treatment (in dashed black).[12] The program effect is measured at the cut point, denoted by 0 in the graph.

More formally, Table 2 provides relatively clear and consistent evidence of negative effects, particularly in math, for various subgroups defined by gender, race, or free and reduced price lunch (FRL) status. Results are reported by post-program year and for various model specifications. The first row of this table provides the first stage estimate of the increase in assignment to the treatment caused by the discontinuity.[13] As expected, there is a strong uptick in treatment probability at the discontinuity, and the F-statistic for the first stage is well above the recommended minimum of 10 (Angrist & Pischke, 2009; Staiger & Stock, 1997).

The estimated treatment effects on test scores are in the following rows. Although the estimates differ somewhat across specifications and are not all statistically significant, all of the coefficients for both math and reading overall and for subgroups defined by gender, race, and SES are negative for both 2013 and 2014. Of note are the consistently large and significant negative effects in math for female, Hispanic, and FRL students in 2014, and the negative effects for Black students in reading in 2014. We can rule out the possibility that these negative findings reflect prior year trends by extending the basic analysis back in time to 2006, as shown in Figure 3. In the subgroup of schools that were open from 2006-2014, we find strong negative effects in the overall composite score in 2014, in math in 2013 and 2014, and in reading in 2014. Importantly, we find no evidence of effects in 2006 through 2010.

To supplement our analysis of how the program affected student test scores in the treated schools, we also explored how it affected student behavior (see bottom part of Table 2). Although one might hope that the program would increase a school's average student attendance, it apparently decreased average attendance by 0.4 to 1.2 percentage points in 2012, though the effect dissipates in later years. At the same time, we find some evidence that the program resulted in a higher rate of student suspensions in 2012, ranging from a 6.5 to 21.6 increase in suspensions per 100 students. In sum, the schools subject to the state's turnaround program exhibit worse or no better student outcomes than comparable untreated schools.

Next, we turn to the student level longitudinal analysis. The sample includes students in schools at various bandwidths from the cut points. Although these students have test scores below the state average, students in schools just above the cut point are similar to students in schools just below the cut point. The columns labeled "all" in Table 3 shows that the program had no observable overall effect on the passing rates of the treated students in either math or reading, where passing is defined as being at level III or IV on the state's four level scale, and the treated students are those who were in treated schools in third grade. This null average effect, however, masks some differential effects by student achievement level. For grade 3 students who were at Level II in math – that is, just below passing – in 2010 we find weak evidence that the turnaround program increased their probability of passing by 10.1 to 21.2 percentage points in 2012, when most of them were in fifth grade. These are matched with a 0.13 to 0.29 SD increase in test scores for this group. The magnitude and precision, though not the direction, of these estimates are sensitive to our choice of bandwidth. Hence this evidence is at best suggestive. Moreover the gains faded as the students continued to progress through school, presumably as many of them moved to middle schools that were not turnaround schools (results not shown). Any initial positive effect for this group of students would be consistent with the view that teachers in the turnaround schools concentrated more effort on students at the borderline of passing than did teachers in other schools.

At the same time, we find consistently large reductions (0.36 to 0.64 SD) in reading scores for those who were in the highest category in 2010. There is no associated drop in passing, likely because these students score well above the passing mark. Recall that we follow students regardless of their 2012 school. Hence the observed decline in the test scores of the highest achievers is consistent either with teachers concentrating less attention on them or on potential negative effects from changing schools, a topic to which we return below.

In sum, the turnaround program did not increase average achievement at either the school level or the student level. Instead it appears to have reduced overall passing rates at the school level. The only group that may have gained from the program was the students who were just below passing in 2010, though these gains do not persist and are not consistent across specifications.

5. Results

The effects on student outcomes, particularly those at the school level, are clearly inconsistent with the goals of the state's turnaround program. With our detailed data on teachers, principals, teacher behavior, and school climate we are in a position to explore possible explanations for the disappointing results. These explanations include the possibility that the program was not fully implemented, that it reduced principal or teacher quality, that it put inappropriate demands on teachers, that it weakened or at least did not improve the school climate, or that the program exacerbated the problems of the low performing schools by increasing their proportions of disadvantaged students. We warn the reader that we are not in a position to draw strong conclusions about the contribution of specific explanations to the overall patterns of student outcomes. Instead, we use the analysis to determine the causal effects of TALAS on various school level variables, which in turn allow us to speculate about why the program did not improve student outcomes. If we do not observe a change in a specific variable, we can effectively rule it out as a causal explanation for the changes in the test scores.

5.1 Effects on principal and teacher turnover

We begin by examining how the TALAS program affected the turnover of principals and teachers. Although the federal government guidelines provided four school turnaround models (transformation, turnaround, restart, or closure), NCDPI officials recognized that it would be difficult for many rural schools to close or to replace 50% of their staff as requited under the turnaround model. As a result about 85% of the TALAS schools, and all of the rural TALAS schools, chose the transformation model, which focused on the removal of the principal but not the removal of staff.

Figure 4 and Table 4 indicate that the program did lead to significantly higher principal turnover. Consistent with the heavy use of the transformation option, we find that school principals left the treated schools at higher rates than in the other schools during 2012, the first full year after the program was implemented.[14] Although policymakers may assume that removing a principal is an appropriate strategy for failing schools, its effectiveness depends on whether the new principals are more effective than the departing principals. We do not have much information on that issue, but Table 4 shows that the program led to a higher proportion of principals with limited experience (less than 3 years), possibly in all three years, but quite clearly and consistently by 2014. These findings from the RD analysis are consistent with descriptive analyses that show higher overall rates of principal departure from the treated schools than from the control schools by 2014 (about 92% vs. 70% from 2010 to 2014). A higher percentage of the replacement principals in the treated schools came from the new principal pool, compared to the control schools which were more likely to hire principals from other schools. If inexperienced principals are less effective than more experienced principals, the decline in principal quality could potentially account for some of the observed decline in student passing rates.

For teachers, we find an uptick in turnover in the year after the increase in principal turnover (see the right part of Figure 4).[15] We cannot say for certain why turnover increased. It could be because teachers waited to experience a full year of the program before changing schools, or because new principals had to wait a year to make staffing changes. We note that several schools mentioned placing low-performing teachers on action plans in 2012, with the intention to remove them if they do not achieve growth. Other schools mention an increase in teacher resignations in 2013 for teachers not meeting principal expectations (Department of Public Instruction, 2014).[16] As reported in Table 4, we find no change in the proportion of inexperienced teachers, so we cannot attribute the fall in student passing rates to an increase in inexperienced teachers. Nonetheless, we note that teacher turnover in a schools can be disruptive to student learning (Ronfeldt, Loeb, & Wyckoff, 2013). We can rule out the possibility that these findings reflect prior year trends by extending the basic analysis back to 2009, as shown in Figure 5. We find no effect in the placebo pre-treatment years, but a large effect in 2012 for principal turnover and in 2013 for teacher turnover.

5.2 Effects on how teachers spend their time

The turnaround and transformation models required several changes to teacher behavior. Under the transformation model, the district must "provide staff with ongoing, high-quality, job-embedded professional development" and "promote the continuous use of student data (such as from formative, interim, and summative assessments) to inform and differentiate instruction in order to meet the academic needs of individual students." Schools must also increase "learning time and create community-oriented schools," with a specific requirement to "provide ongoing mechanisms for family and community engagement" (Race to the Top, 2014) Using the teacher survey data on time use, we examine the extent to which the program affected how teachers spent their time in schools. We group these activities into four categories: (1) activities that may improve teachers (i.e., professional development, individual planning time, collaborative planning time, and utilizing the results of assessments), (2) greater administrative burdens (i.e., supervisory duties, required committee/staff meetings, and paperwork), (3) attention to community issues and student problems (i.e., communicating with parents/community members and addressing student discipline), and (4) focusing on tests (i.e., preparation and delivery of federal, state, and local tests). Several of these activities are specifically identified as required as part of the transformation and turnaround models, but others are not. We predict hours spent on these activities in 2012 and 2014, though some caution may be necessary for 2014 given the high teacher turnover in treated schools in 2013.

Figure 6 illustrates the changes for the group of activities involving teacher improvement. Among the activities portrayed in Figure 6, TALAS appears to have had a large positive effect on professional development and collaborative planning. The formal statistical analyses of the patterns for all the teacher activities are shown in Table 5. The most consistent 2012 findings emerge for professional development, supervisory duties, required committee or staff meetings, and required paperwork, each of which increase as a result of the program. Professional development was meant as a key component of the TALAS program, so its increase is expected. Although community involvement was meant to be part of the TALAS program, the program apparently had little effect on the amount of time teachers devote to community, parents, or student conduct in 2012, but communicating with parents and the community did increase by 2014. Teachers also spend more time delivering assessments in treated schools by 2014, but they did not change the time they spent using the results of these assessments.

It is difficult to predict the contributions of these changes to changes in student outcomes. More time in professional development could be positive in the long run provided the development is high quality, but it could take time away from teaching in the short run. In the short run, the additional time for collaborative planning could well be productive. More time in required meetings and filling out paperwork, however, is not likely to be productive as it takes time way from instruction. Additional insight into these changes emerges from teachers' perceptions of their working conditions, to which we now turn.

5.3 Effects on teachers' perceptions of their school climate

Table 6 reports effects on teachers' perceptions of school climate based on factors calculated from the working conditions survey. Positive numbers indicate increases in satisfaction in treated schools. Despite the fact that turnaround models emphasize school leadership and that school leaders changed in many schools, the TALAS program apparently had no effect on teachers' perceptions of the quality of their schools' leadership, perhaps because many of the new principals were inexperienced. Nor did it have much effect on teachers' perceptions of the quality of other activities including professional development or community involvement. Some hints of dissatisfaction with facilities and resources emerge in the 2012 survey, along with some concerns about time pressures in the 2014 survey. We remind the reader that we are not simply looking at survey results, but rather at estimates of how the TALAS program affected the responses.

Combining these findings related to teachers' perceptions of their school's climate with those related to their activities and use of time, we conclude that the TALAS program generated few significant changes for teachers that would be consistent with an academically more productive environment in the schools, at least in the short run. Conceivably more professional development or collaborative planning could help teachers, but the clearest picture that emerges in the post-turnaround environment is one in which teachers have heavier administrative burdens, more paperwork, and a sense that they have fewer resources.

5.4 Effects on the concentration of disadvantaged students

The TALAS program focuses attention on schools, but individual schools could be serving a changing mix of students during the study period. Hence, a final possibility is that the decline in the school-level performance in the treated schools may be caused by the flight of high-achieving students and an increasing concentration of low-achieving students. If assignment to turnaround status stigmatizes a school or if parents do not like the changes in the schools, more advantaged students might move to other schools, leaving greater concentrations of lower-scoring disadvantaged students behind.

We find evidence that TALAS led to such differential movement of students. Figure 7 displays an RD analysis that focuses on students who were third or sixth graders in schools +/-16 percentage points from the cut point in 2010. The Y axis displays the proportion of students who remain in the same school through 2012, when they would likely be fifth or eighth graders (though we retain students who failed or skipped a grade in the analysis). We find that the chance that FRL students change schools is fairly constant across the cut point. However non-FRL students are much less likely to remain in the same school if they are in a school assigned to treatment in 2010, relative to the FRL students (*p-value*=0.009). In other words, more affluent students from treated schools are more likely to attend a different school two years later.

With 83 percent of the students in our elementary school sample eligible for free or reduced price lunch, the differential movement of low and high income students may not translate to large overall effects at the school level. Table 7, which examines the effect of TALAS at the school level, however, provides some evidence that the program did increase the share of FRL students in the treated schools. For all years and across all methods, the estimated coefficients indicate an increase in the share of the percent of FRL students in the treated schools. There is no effect for the percentages of black or Hispanic students.

In sum, this evidence of a higher proportion of students on free and reduced price lunch in treatment schools after the TALAS program was implemented may account for some of the decline in student outcomes at the school level. Analysis of student movement is important in that it highlights that school outcomes depend not only on a school's practices but also on the mix of students in the school. In this case, the movement of students exacerbates the challenge of transforming low-performing schools into higher-performing schools. Given the small magnitude of the effects on the proportion of FRL students, however, one should not attribute the entire decline in school level test scores to the changing mix of students.

6. Robustness Checks and Alternative Explanations

An RD design relies on the assumption that assignment is "as good as random" around the cutoff point, or, alternatively, that we have specified the correct functional form. We have already reported several findings relevant to the validity of the assumptions that underlie our analysis, specifically finding that schools did not manipulate the assignment variable and that baseline characteristics are balanced at the cutoff. Van der Klaauw (2008) recommends using outcome data from a period before the program was put into place as a falsification or placebo test. With minimal marginally significant exceptions, we found no such placebo discontinuities, indicating that the effect came from the program itself (see Table 1, the first column of Tables 5-6, and Figures 3 and 5). In addition, we used several models at different bandwidths to increase our confidence in our estimates.

Finally, other programs simultaneously occurred in North Carolina during this time and may have affected our estimates if their uptake was discontinuous at the TALAS cutoff point. For instance, NCDPI's Federal Programs division operates programs required by the Elementary and Secondary Education Act (Department of Public Instruction, 2015). Interviews with NCDPI staff indicate that the Federal Programs and turnaround divisions are distinct, with Federal Programs focusing on monitoring and TALAS on coaching, but some of the Federal Programs projects target schools similar to our TALAS schools. In analysis shown in the Appendix, we check to make sure there is no jump in the probability of assignment to one of these programs at our cutoff, which would violate the exclusion restriction. We find no evidence of such a jump, which gives us confidence in our estimates of the effects of the TALAS program. However, the appearance of these other programs cautions against making causal claims about schools well away from the cutoff.

7. Conclusion

We find very little evidence that North Carolina's TALAS program, which was funded by federal Race to the Top money and designed to turn around the state's lowest performing schools, had the intended positive effects for elementary and middle schools near the cut point for eligibility. Hence, our results provide strong causal evidence against expanding the TALAS program at the margin. We cannot make strong conclusions about the effectiveness of the program for schools away from the margin, as schools well below the cut point were subject to other programs. However, if the program did not work well for schools near the eligibility cutoff, it seems unlikely that it would work for those well below that point.

Although the ultimate goal of the program was to improve student test scores, it instead led to a drop in school-wide passing rates in math (especially for female and Hispanic students) and in reading (especially for Black students). Among students who experienced the program in the first full treatment year, the program may have helped those on the borderline of passing in math, but it decreased the scores of the highest-achieving students in reading. In addition, we provide some limited evidence that the program led to an increase in the proportion of disadvantaged students in the treated schools.

Our unique statewide data set based on the state's biannual Teacher Working Conditions Survey allowed us to open the black box to examine how teacher activities change under a turnaround regimen. We find that substantial change occurred in the treated schools. As required by the program, the schools brought in new principals and increased the time teachers devoted to professional development. But the program also increased administrative burdens and distracted teachers, potentially reducing the time available for instruction. Teachers became less satisfied with the time and other resources they had available and their turnover increased after the first full year of implementation. While strong leadership and changes to instructional practices may be important in general for turning around low-performing schools, North Carolina's mixture of principal replacement and teacher professional development were apparently not sufficient to generate the positive changes in instructional practices or transformational leadership needed to raise student achievement in those schools, and indeed appears to have reduced it.

Our analysis is necessarily limited to relatively short run effects, namely effects in 2012 (the first year after the program was fully implemented), 2013, and 2014. Hence, we cannot rule out the possibility that more positive effects may emerge over time. A report on the North Carolina program on which TALAS is based clearly emphasized the need for continuity (Thomson et al., 2011). Although researchers should continue to follow-up with these schools, the short-term nature of Race to the Top funding could make program sustainability difficult (Anrig, 2015).

At the same time, we are not optimistic about the program's future success in part because it may be focusing on the wrong objects. To the extent that the failure of low performing schools reflects the challenges that disadvantaged students bring to the classroom, and not simply poor leadership or instruction, more attention to those challenges may be necessary in the form, for example, of health clinics, counselors, or mental health specialists.[17] Moreover, disadvantaged students clearly need effective teachers and within-school structures of academic and social support to succeed. We found little evidence that North Carolina's turnaround program led to changes of this type in the state's lowest performing schools, and hence it is not surprising that the program failed to realize its goals. One potential lesson from this North Carolina experience is that turning around low-performing schools is difficult, and that, while changes in leadership and other short term changes may often be necessary for such change, they are far from sufficient to address the deep long term challenges that such schools face.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444. http://doi.org/10.2307/2291629

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Anrig, G. (2015). *Lessons from School Improvement Grants That Worked* (Issue Brief) (pp. 1–21). New York, N.Y.: The Century Foundation. Retrieved from http://www.tcf.org/bookstore/detail/lessons-from-school-improvement-grants-that-worked

Boyd, D., Lankford, H., & Wyckoff, J. (2007). Increasing the effectiveness of teachers in low-performing schools. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (1st ed., pp. 612–630). Routledge.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (Working Paper No. 17803). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17803

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, *24*(4), 409–429. http://doi.org/10.1037/h0027982

Clark, D. (2009). The performance and competitive efffects of school autonomy. *Journal of Political Economy*, *117*(4), 745–783. http://doi.org/10.1086/605604

Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance* (Working Paper No. 38). New York, N.Y. Retrieved from http://eric.ed.gov/?id=ED509693

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682. http://doi.org/10.1016/j.econedurev.2007.10.002

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655–681.

Conaty, J. C. (2011, January 31). Race to the Top Amendment Approval. Retrieved from http://www2.ed.gov/programs/racetothetop/amendments/north-carolina-2.pdf

Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis*, *25*(4), 375–395. http://doi.org/10.3102/01623737025004375

Cowen, J. M., Butler, J. S., Fowles, J., Streams, M. E., & Toma, E. F. (2012). Teacher retention in Appalachian schools: Evidence from Kentucky. *Economics of Education Review*, *31*(4), 431–441. http://doi.org/10.1016/j.econedurev.2011.12.005

Dee, T. S. (2012). *School turnarounds: Evidence from the 2009 stimulus* (Working Paper No. 17990). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17990

Department of Public Instruction. (2011). NC Middle and High School Turnaround. Retrieved September 23, 2014, from http://www.ncpublicschools.org/schooltransformation/turnaround/

Department of Public Instruction. (2013a). *APR Report 2013 on Status of RttT TALAS Implementation by School* (RttT TALAS Report). Raleigh, NC: North Carolina Department of Public Instruction. Retrieved from https://www.rtt-apr.us/state/north-carolina/2012-2013/talas

Department of Public Instruction. (2013b). *Race to the Top North Carolina Report Year 2: 2011-12*. Washington, D.C.: U.S. Department of Education. Retrieved from http://www2.ed.gov/programs/racetothetop/performance/north-carolina-year-2.pdf

Department of Public Instruction. (2014). *Race to the Top North Carolina Report Year 3: 2012-13*. Washington, D.C.: U.S. Department of Education. Retrieved from https://www.rtt-apr.us/sites/default/files/state_year_pdfs/north-carolina/2012-2013/state-specific-summary-report__north-carolina__2012-2013.pdf

Department of Public Instruction. (2015). Federal Program Monitoring. Retrieved August 6, 2015, from http://www.ncpublicschools.org/program-monitoring/

Duffrin, E. (2012). North Carolina. In U. Boser, *Race to the Top: What Have We Learned So Far?* (pp. 47–50). Washington, D.C.: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/issues/2012/03/pdf/rtt_states.pdf

Fan, J., & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. CRC Press.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, M., … Sztenjnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE No. 2008-4030). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/quickreviewsum.aspx?sid=92

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., … Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE No. 2011-4024). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/quickreviewsum.aspx?sid=92

Goldberger, A. S. (1972). *Selection bias in evaluating treatment effects: The case of interaction*. Unublished manuscript, Madison, WI.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209. http://doi.org/10.1111/1468-0262.00183

Hansen, M. (2013). *Investigating the role of human resources in school turnaround: A decomposition of improving schools in two states* (Working Paper No. 89). National Center for Analysis of Longitudinal Date in Education Research. Retrieved from www. caldercenter. org/publications/upload/wp89. pdf

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39*(2), 326–354. http://doi.org/10.3368/jhr.XXXIX.2.326

Henry, G. T., Campbell, S. L., Thompson, C. L., & Townsend, L. W. (2014). *Evaluation of district and school transformation school-level coaching and professional development activities* (pp. 1–56). Chapel Hill, NC: Consortium for Educational Research and Evaluation. Retrieved from http://cerenc.org/wp-content/uploads/2014/10/DST-School-Level-Coaching-10-2-14.pdf

Herman, R., Dawson, P., Dee, T. S., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *Turning Around Chronically Low-Performing Schools: What Works Clearinghouse* (NCEE No. #2008-4020). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/practiceguide.aspx?sid=7

Imbens, G. W., & Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 933–959. http://doi.org/10.1093/restud/rdr043

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, *27*(2), 213–256. http://doi.org/10.1086/599334

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*(2), 235–261. http://doi.org/10.3102/0162373711398128

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. http://doi.org/10.1257/jel.48.2.281

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44–70. http://doi.org/10.1207/s15327930pje8003_4

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370–397. http://doi.org/10.1177/0013161X03253412

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714. http://doi.org/10.1016/j.jeconom.2007.05.005

Moore Johnson, S., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1–39.

Nichols, A. (2011). rd 2.0: Revised Stata module for regression discontinuity estimation (Version 2.0) [Stata]. Washington, D.C.: Urban Institute. Retrieved from http://ideas.repec.org/c/boc/bocode/s456888.html

Race to the Top. (2014). Definitions. Retrieved December 10, 2014, from https://www.rtt-apr.us/about-apr

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36. http://doi.org/10.3102/0002831212463813

RttT Application. (2010). *Race to the Top Application*. Raleigh, NC: Office of the Governor & NC Department of Public Instruction.

Schochet, P. Z. (2009). Statistical power for regression discontinuity designs in education evaluations. *Journal of Educational and Behavioral Statistics*, *34*(2), 238–266. http://doi.org/10.3102/1076998609332748

Shadish, W. R., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D. (2011). A randomized experiment comparing random and cutoff-based assignment. *Psychological Methods*, *16*(2), 179–191. http://doi.org/10.1037/a0023345

Sipple, J. W., & Brent, B. O. (2007). Challenges and strategies associated with rural school settings. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (1st ed., pp. 612–630). Routledge.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586. http://doi.org/10.2307/2171753

Tang, Y., Cook, T. D., & Kisbu-Sakarya, Y. (2014). *Reducing bias and increasing precision by adding either a pretest measure of the study outcome or a nonequivalent comparison group to the basic regression discontinuity design*. In preparation, Northwestern University.

Thomson, C. L., Brown, K. M., Townsend, L. W., Henry, G. T., & Fortner, K. C. (2011). *Turning around North Carolina's lowest achieving schools (2006-2010)*. Chapel Hill, NC: Consortium for Educational Research and Evaluation - North Carolina. Retrieved from publicpolicy.unc.edu/files/2014/02/DST_1st-year-Report_FINAL_12-05-2011.pdf

Van Der Klaauw, W. (2008). Regression-discontinuity analysis: A survey of recent developments in economics. *Labour*, *22*(2), 219–245. http://doi.org/10.1111/j.1467-9914.2008.00419.x

Whalen, A. (2011, December 27). Race to the Top Amendment Approval. Retrieved from http://www2.ed.gov/programs/racetothetop/amendments/north-carolina-9.pdf

Wing, C., & Cook, T. D. (2013). Strengthening the regression discontinuity design: A within-study comparison. *Journal of Policy Analysis and Management*, *32*(4), 853–877. http://doi.org/10.1002/pam.21721

Appendix

This appendix describes the methods used to create our school climate construct and potential discontinuities in simultaneously-occurring programs.

School Climate Constructs

This section provides details on North Carolina's biannual Teacher Working Conditions Survey and our factor analysis strategy. Teachers answered 83 questions about school climate that appeared on the 2010, 2012, and 2014 versions of the survey. We used the *factor* program in STATA 12 to break these questions into related factor constructs (using principal factor analysis). We took the factors with Eigen values above one to create seven constructs: leadership, instructional practices, professional development, community involvement, student conduct, facilities and resources, and time use. We used the variable weighting from the 2010 factor analysis on 2012 and 2014 data to create 2012 and 2014 factors, respectively.

Table A1 displays the survey wording, the top factor for each question as indicated by the factor analysis, and a splined linear estimate for the effect of treatment on the factor in 2012 and 2014 for our two main bandwidths. Each construct may have weight in multiple constructs; the table displays the main factor component for each question. Using this primary category, the constructs have the following Cronbach's alphas: leadership (0.991), instructional practices (0.900), (professional development (0.976), community involvement (0.961), student conduct (0.950), facilities and resources (0.921), and time use (0.921).

Within Instructional Practices, treated teachers are particularly dissatisfied with local assessment data being available in time to impact instructional practices in 2014. Within the Time construct, treated teachers are particularly dissatisfied with being able to focus on students with minimal interruptions (in 2014), the amount of instructional time to meet all students' needs (in 2014), and being protected from duties that interfere with their essential role of educating students (in 2012 and 2014).

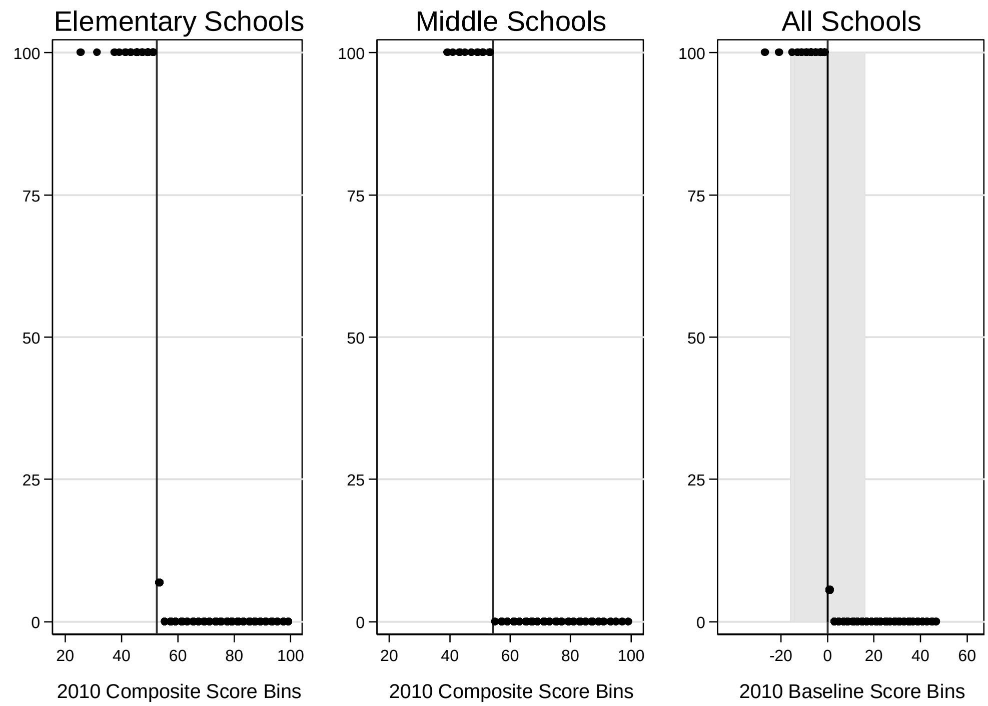
Discontinuities in Simultaneous Programs

There are three ESEA school distinctions: Reward, Focus, and Priority. Reward Schools are recognized as either high-achieving or high-growth with banners and public recognition. NCDPI must also recognize 5% of Title I schools as Priority and 10% as Focus Schools, at which point local school districts must provide various programs to students. The worry with these programs might be that recognition by DPI or programs run by the district might overlap with the work at TALAS schools.

Because we estimate the effect of the TALAS program at the cutoff point only, there would have to be a difference in the ESEA program assignment *at the 2010 cutoff.* Importantly, TALAS and ESEA schools do not have the same assignment mechanism. Assignment to an ESEA distinction can be dependent on growth or absolute scores, with the 2011 school year as a baseline. Because scores are somewhat random from year-to-year, and because TALAS schools are selected only on absolute scores from 2010, we do not expect a strong relationship between our discontinuity point and assignment to these programs. Indeed, this is the case, with no relationship between these programs at the cutoff point (see Figure A1). The assignments largely match expectations, with higher-achieving schools more likely to receive Reward distinction and lower-achieving schools more likely to be labeled Priority. However, the probability of assignment to these distinctions is about equal just above and below the cutoff point. This gives us confidence about our estimate as a LATE.

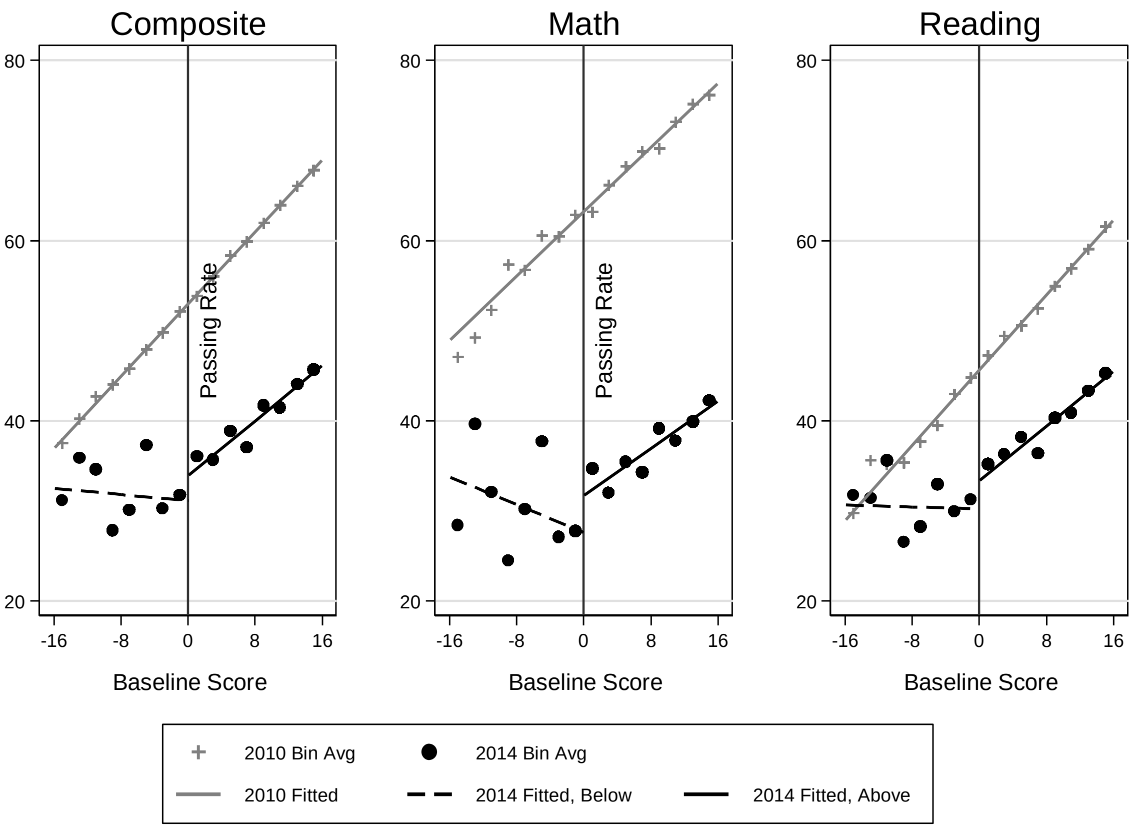
Figures

Figure 1: Treatment Uptake by School Type



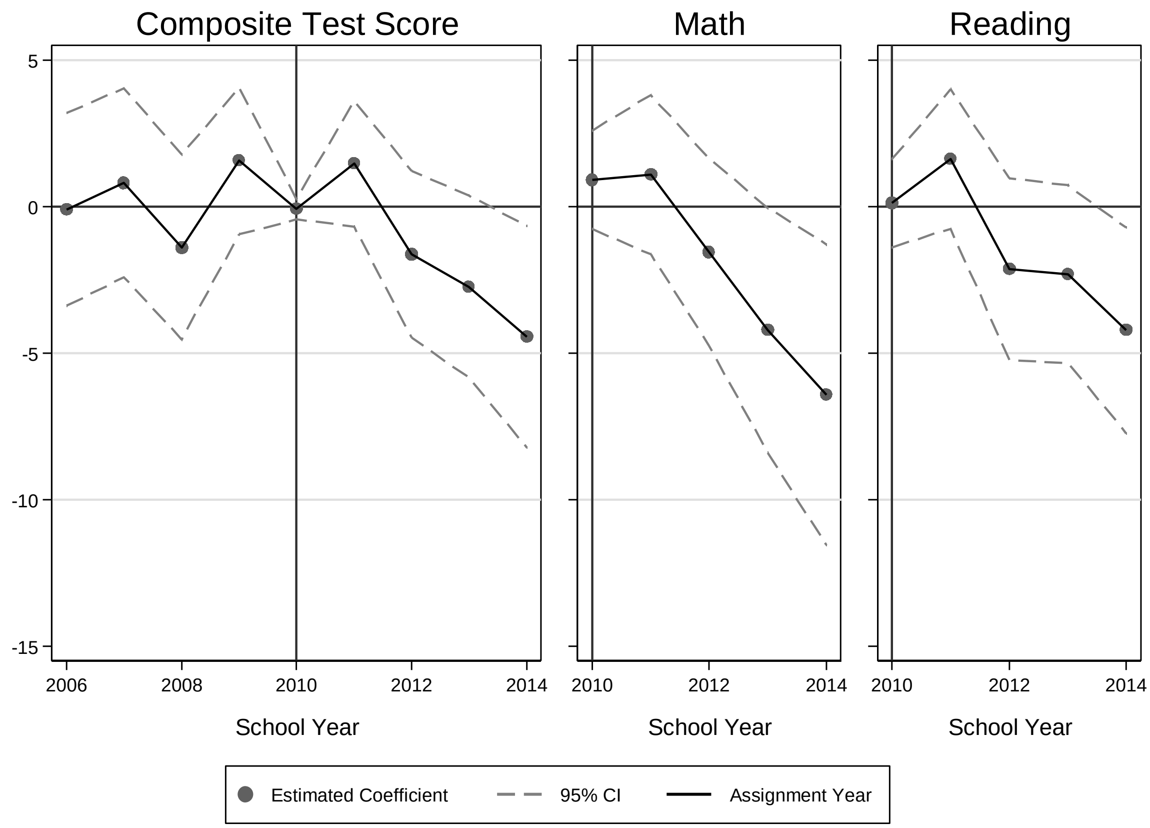
Note: Charts display the average uptake within 2.0 percentage point bins. Line indicates 2010 composite score cutoff. Grayed area indicates +/-16% from baseline cutoff.

Figure 2: 2014 Composite, Math, and Reading Scores



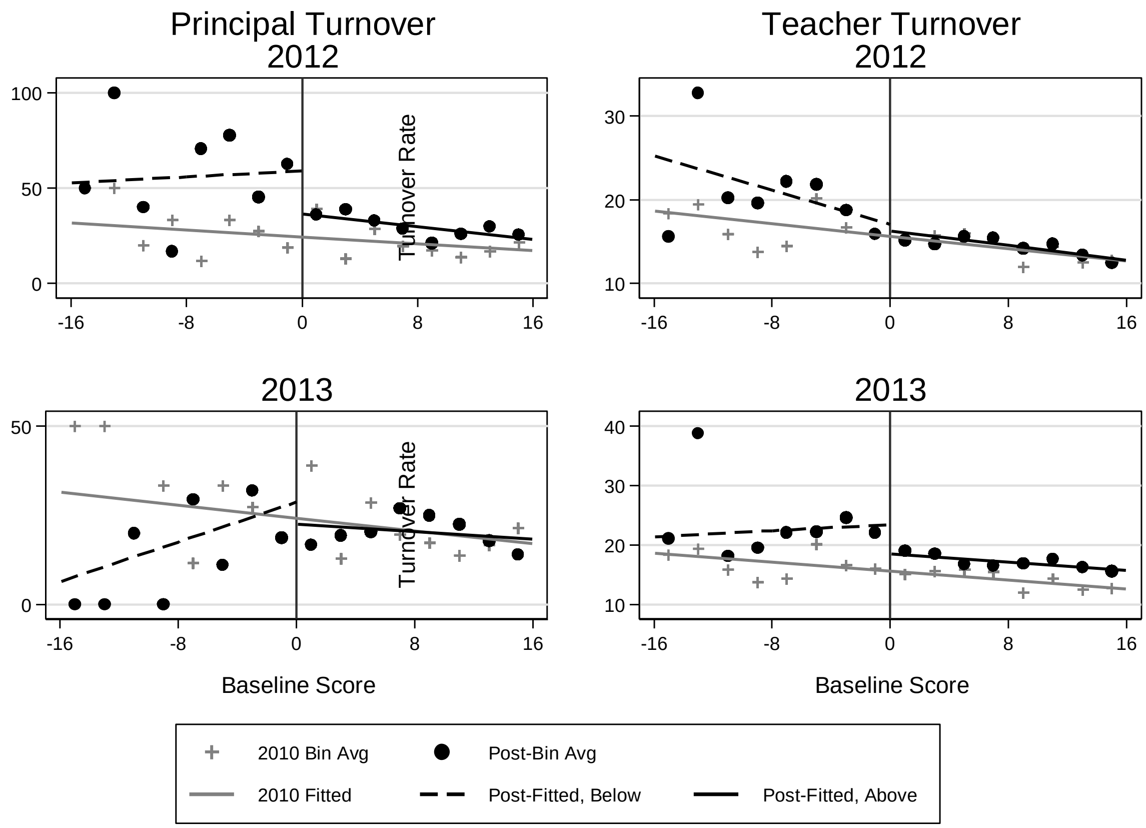
Note: Estimates of outcomes in 2010 and 2014 within +/-16% using our linear spline model with no additional controls (N=518 schools). Untreated post-period segment not constrained to be parallel with pre-period segments. All scores dropped from 2010 to 2014 due to a change in testing. Displayed bin width=2-percentage points.

Figure 3: Test Results by Year



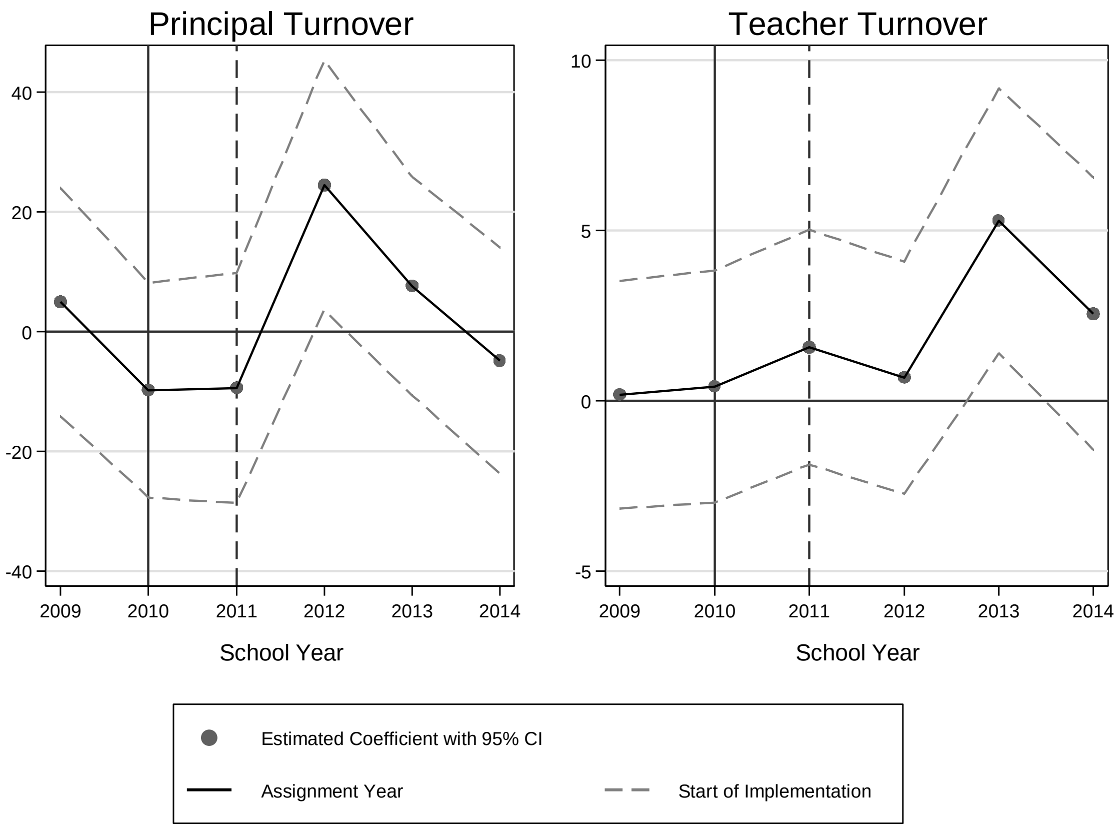
Note: Based on a separate +/-16% linear spline estimate with no additional controls for each year. Only includes schools that appear in all years 2006-2014 (N=493 schools per year) to avoid compositional effects from schools that closed or opened over the period.

Figure 4: 2012 and 2013 Principal and Teacher Turnover



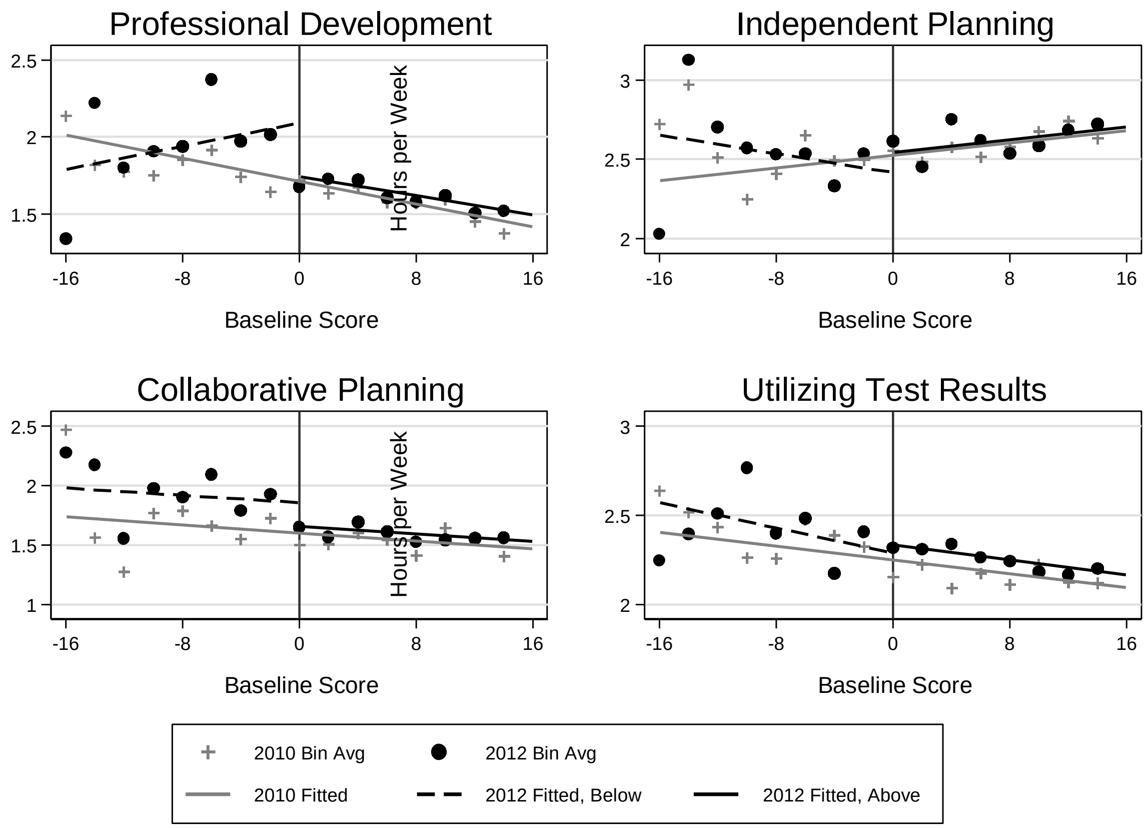
Note: Estimates of outcomes in 2010 and 2014 within +/-16% using our linear spline model with no additional controls (N=518 schools). Untreated post-period segment not constrained to be parallel with pre-period segments. Displayed bin width=2-percentage points.

Figure 5: Staff Turnover by Year



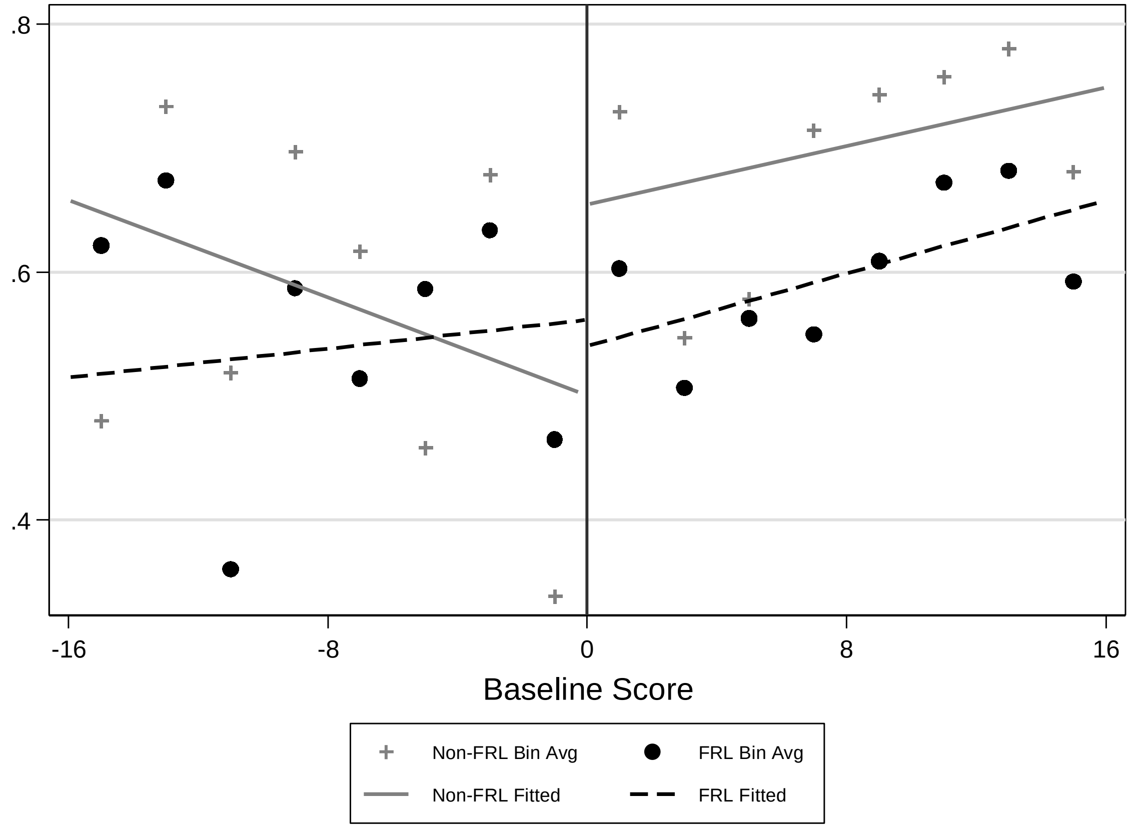
Note: Based on a separate +/-16% linear spline estimate with no additional controls for each year. Only includes schools that appear in all years 2009-2014 (N=512 schools per year) to avoid compositional effects from schools that closed or opened over the period.

Figure 6: 2012 Hours Spend on Activities per Week



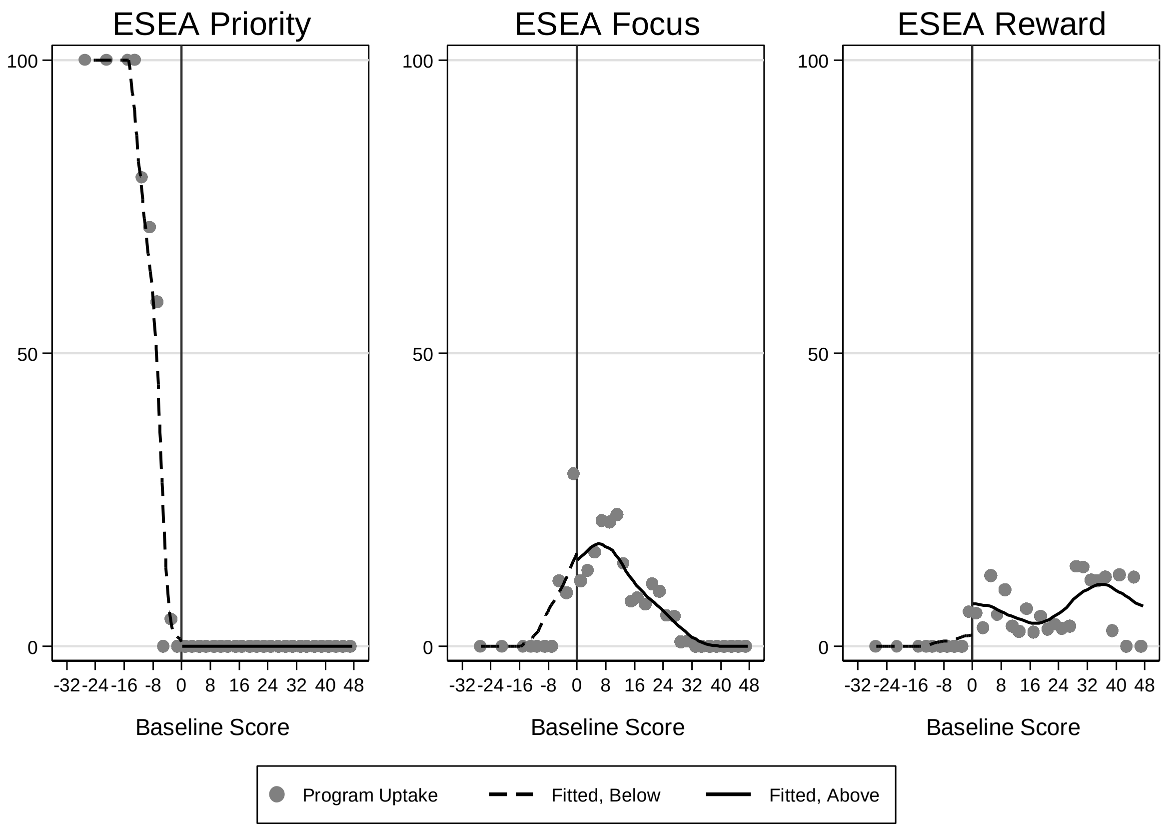
Note: Estimates of outcomes in 2010 and 2014 within +/-16% using our linear spline model with no additional controls (N=518 schools).. Untreated post-period segment not constrained to be parallel with pre-period segments. Displayed bin width=2-percentage points.

Figure 7: Student-Level Movement



Note: Estimates of probability of remaining in the same school from 2010 to 2012 for students who were in third or sixth grade in 2010. Analysis conducted at the student level within +/-16% of the 2010 schools using our linear spline model with no additional controls. Displayed bin width=2-percentage points.

Figure A1: Uptake of ESEA Reward/Priority/Focus Schools (Fraction)



Note: Nonparametric estimates based on 100% IK bandwidth. ESEA Priority uses 50% of IK bandwidth because 100% bandwidth predicts a negative number of schools at the cut point. Displayed bin width=2-percentage points.

Tables

Table 1: Comparison of 2010 Baseline Characteristics Above and Below the Cutoff Value

	Panel A: Average Value (+/-16%)			Panel B: Estimated Value at Cutoff[1]		
2010 Values	Below Cutoff (-16% to 0%)	Above Cutoff (0 to 16%)	P-value of Difference	Below Cutoff	Above Cutoff	P-value of Difference
Assignment Score	-5.158	9.285	0.000 ***	0.000	0.000	N/A
	(0.412)	(0.212)		(0.000)	(0.000)	
Percent FRL in School	86.410	75.269	0.000 ***	83.746	86.122	0.331
	(1.253)	(0.602)		(2.444)	(1.149)	
Percent Black in School	64.886	46.888	0.000 ***	59.557	59.201	0.946
	(2.718)	(1.033)		(5.298)	(2.278)	
Percent Hispanic in School	16.001	16.411	0.819	17.728	16.404	0.673
	(1.825)	(0.685)		(3.133)	(1.540)	
Student Daily Attendance	94.478	94.861	0.002 **	94.872	94.497	0.147
	(0.121)	(0.048)		(0.259)	(0.117)	
Short Term Suspensions	32.266	20.638	0.000 ***	27.476	27.560	0.990
	(3.226)	(1.057)		(6.433)	(2.569)	
1-Year Principal Turnover	24.051	20.501	0.477	20.301	27.466	0.467
	(4.839)	(1.929)		(9.852)	(4.851)	
1-Year Teacher Turnover	16.278	13.952	0.013 *	16.715	16.370	0.860
	(1.046)	(0.347)		(1.952)	(0.882)	
Teachers w/ 0-3 Yrs. Exp.	25.467	23.640	0.148	24.720	26.462	0.423
	(1.089)	(0.498)		(2.175)	(1.049)	
N	79	439				

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Panel B based on a parametric RD with a linear spline function for schools +/-16% from the cutoff with no additional control variables (X_s). Robust standard errors in parentheses.

Table 2: School-Level Math, Reading, and Behavioral Outcomes; Estimates by Method, Bandwidth, and Year

	2012			2013			2014		
	NP(1)	2SLS(2)		NP(1)	2SLS(2)		NP(1)	2SLS(2)	
Bandwidth	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%
First Stage	0.977***	0.977***	0.912***	0.872***	0.946***	0.912***	0.885***	0.945***	0.911***
	(0.016)	(0.016)	(0.051)	(0.074)	(0.034)	(0.051)	(0.067)	(0.035)	(0.052)
F-Statistic	N/A	51993.67	30789.36	N/A	51524.71	30789.36	N/A	49085.27	29946.69
End-of-Grade Math Passing Rates									
Overall	1.125	-1.521	0.171	-5.267+	-3.299	-2.465	-6.094	-5.108+	-3.655
	(2.263)	(1.865)	(2.185)	(2.948)	(2.117)	(2.476)	(3.763)	(2.677)	(3.095)
Male Students	0.495	-2.186	-1.024	-6.064	-2.805	-1.828	-5.370	-4.402	-2.705
	(2.332)	(1.980)	(2.267)	(3.952)	(2.433)	(2.857)	(3.817)	(2.756)	(3.262)
Female Students	0.388	-0.810	1.248	-6.127*	-4.021*	-3.358	-6.461+	-5.428*	-4.051
	(2.324)	(2.001)	(2.450)	(2.625)	(2.004)	(2.338)	(3.898)	(2.761)	(3.183)
Black Students(3)	0.293	-0.556	0.059	-4.831+	-3.943*	-2.441	-1.591	-3.279	-1.239
	(2.794)	(2.121)	(2.524)	(2.826)	(1.722)	(2.014)	(3.448)	(2.591)	(2.977)
Hispanic Students(3)	0.576	0.704	0.828	-6.691+	-5.185	-5.777	-8.319+	-6.719+	-7.156+
	(3.454)	(2.518)	(2.947)	(3.568)	(3.245)	(3.548)	(4.676)	(3.495)	(4.095)
FRL Students	2.148	-0.922	0.810	-2.726	-3.176	-2.264	-4.757	-4.675+	-2.995
	(2.929)	(1.846)	(2.185)	(2.756)	(2.006)	(2.339)	(3.817)	(2.632)	(3.003)
End-of-Grade Reading Passing Rates									
Overall	-0.486	-1.898	-0.216	-5.464*	-1.802	-2.517	-3.440	-3.225+	-2.912
	(2.113)	(1.465)	(1.819)	(2.678)	(1.488)	(1.873)	(2.568)	(1.860)	(2.294)
Male Students	-1.976	-2.665+	-1.695	-8.163*	-2.964+	-3.795+	-3.764	-3.394+	-2.735
	(2.721)	(1.506)	(1.888)	(3.565)	(1.706)	(2.107)	(2.994)	(2.061)	(2.570)
Female Students	0.103	-1.444	1.041	-3.595	-0.887	-1.428	-3.342	-3.001	-3.028
	(2.461)	(1.776)	(2.205)	(2.239)	(1.485)	(1.906)	(2.401)	(1.904)	(2.322)
Black Students(3)	-0.372	-2.018	-0.656	-2.555	-1.809	-2.740+	-2.757	-3.799*	-3.430+
	(2.079)	(1.742)	(2.098)	(1.895)	(1.260)	(1.593)	(2.354)	(1.675)	(2.061)
Hispanic Students(3)	-2.413	-2.749	-1.885	-5.421	-5.340*	-6.463*	-1.555	-3.643	-4.575
	(3.927)	(2.639)	(3.186)	(3.585)	(2.417)	(2.748)	(3.825)	(3.003)	(3.198)
FRL Students	0.476	-1.078	0.615	-2.695	-1.513	-2.354	-0.960	-2.218	-1.794
	(2.295)	(1.421)	(1.740)	(1.960)	(1.332)	(1.663)	(2.706)	(1.740)	(2.141)
Behavioral Outcomes									
Attendance	-1.248**	-0.959*c	-0.394+	-0.685+	0.269q	0.215	0.174	0.173	0.835
	(0.418)	(0.376)	(0.211)	(0.367)	-0.259	(0.219)	(0.953)	(0.478)	(0.574)
Suspensions (per 100 Students)	21.580*	13.672+q	6.473	14.238+	8.821q	3.549	25.924**	4.574	4.601
	(9.500)	(7.276)	(5.400)	(8.029)	-7.079	(5.804)	(9.435)	(4.659)	(5.561)
N	1,753	518	294	1,753	518	294	1,753	518	294
Controls for 2010 baseline composite?	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls for 2010 outcome & school level?	NO	YES	YES	NO	YES	YES	NO	YES	YES

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Nonparametric bandwidths calculated from Imbens and Kalyanaraman (2011).

(2) Linear spline equation used in parametric 2SLS models unless otherwise noted; *q* =quadratic equation used; *c*= cubic equation used.

Table 3: Individual-Level Math & Reading Outcomes; Average Test Scores and Estimated Treatment Effects by Student Baseline Performance Level and Subject, Based on 2SLS Model

		Math(1)					Reading(1)			
Subgroup (based on 2010 Score):		All	Level I	Level II	Level III	Level IV	All	Level I	Level II	Level III	Level IV
2012 Passing Rates											
+/- 16%(2)		-0.000	5.917	10.129	0.041	-1.506	-3.451	0.465	-4.394	-1.221	-3.184
		(2.510)	(19.445)	(8.141)	(2.526)	(1.950)	(3.108)	(5.397)	(7.568)	(3.457)	(3.240)
	N	23398	1143	5410	13620	3225	23277	5988	5369	9645	2275
+/- 10%(2)		4.569	-7.277	20.535*	4.206	-1.638	-0.961	5.499	-2.971	1.154	-4.739
		(2.809)	(20.689)	(9.898)	(2.791)	(2.026)	(3.749)	(6.336)	(9.122)	(4.158)	(4.622)
	N	12887	755	3348	7377	1407	12822	3737	3057	5016	1012
+/- 5%(2)		-0.023	7.947	21.188	-1.172	-1.438	-7.035	10.827	-13.662	-5.313	-9.205
		(4.183)	(31.409)	(13.190)	(4.505)	(1.458)	(6.190)	(9.813)	(13.938)	(6.947)	(7.794)
	N	5639	346	1576	3152	565	5610	1720	1361	2130	399
2012 Standardized Scores											
+/- 16%(2)		0.005	-0.423	0.155	0.008	0.025	-0.016	0.047	0.015	0.001	-0.393*
		(0.069)	(0.373)	(0.109)	(0.071)	(0.147)	(0.049)	(0.099)	(0.086)	(0.057)	(0.175)
	N	23398	1143	5410	13620	3225	23277	5988	5369	9645	2275
+/- 10%(2)		0.086	-0.397	0.289*	0.083	0.121	0.025	0.177	0.083	0.017	-0.356+
		(0.076)	(0.430)	(0.135)	(0.076)	(0.166)	(0.061)	(0.124)	(0.103)	(0.070)	(0.190)
	N	12887	755	3348	7377	1407	12822	3737	3057	5016	1012
+/- 5%(2)		-0.035	0.650	0.127	-0.052	0.179	-0.130	0.305	-0.172	-0.157	-0.641*
		(0.125)	(0.696)	(0.175)	(0.130)	(0.246)	(0.105)	(0.205)	(0.177)	(0.119)	(0.273)
	N	5639	346	1576	3152	565	5610	1720	1361	2130	399

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Columns split into all students from 2010 and separate analyses by 2010 category. Level I and II represent failing ratings.
(2) Analysis uses linear 2SLS models for students who were in treated and untreated schools within the given cutoff in the baseline year. All models control for the school-level baseline composite score, student-level baseline math scores, student-level baseline reading scores, and interactions between these continuous variables, an indicator for being below the assignment score (creating a spline), and the baseline outcome level (to allow for different relationships in the data for different levels of ability). The analysis clusters standard errors for the student's 2010 school. If anything, results are stronger without controlling for both tests; we include both tests to be conservative.

Table 4: Principal and Teacher Turnover; Estimates by Method and Year

	2012			2013			2014		
	NP(1)	2SLS(2)		NP(1)	2SLS(2)		NP(1)	2SLS(2)	
Bandwidth	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%
1-Year Principal Turnover	21.986+	23.129*	18.589	9.993	9.096^{q}	12.766	-5.312	-4.687	-3.464
	(12.138)	(11.166)	(13.748)	(10.803)	(16.186)	(11.260)	(11.055)	(9.917)	(12.061)
Principals with 0-3 Years of Exp.(4)	-0.738	-2.406	28.145^{q}	15.812	23.306*	24.394+	31.589*	27.707*	32.437*
	(11.961)	(11.433)	(20.093)	(14.060)	(11.010)	(13.609)	(14.022)	(11.169)	(13.740)
1-Year Teacher Turnover	1.104	1.037	0.322	3.324	5.292**	5.377*	2.688	2.341	2.810
	(3.024)	(2.227)	(2.617)	(2.585)	(1.771)	(2.181)	(2.568)	(2.399)	(3.000)
Teachers with 0-3 Years of Exp.	2.748	0.021	-0.124	2.708	0.857^{c}	1.821	1.729	1.627	3.701
	(3.597)	(2.490)	(2.983)	(3.520)	(5.484)	(3.106)	(3.841)	(2.732)	(3.097)
N	1753	518	294	1753	518	294	1753	518	294
Controls for 2010 baseline composite?	YES	YES	NO	YES	YES	YES	YES	YES	NO
Controls for school level?	NO	YES	NO	NO	YES	YES	NO	YES	NO

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Nonparametric bandwidths calculated from Imbens and Kalyanaraman (2011).

(2) Linear spline equation used in parametric 2SLS models unless otherwise noted; *q* =quadratic equation used; *c*= cubic equation used.

Table 5: Teacher Time Use; Estimates by Method, Bandwidth and Year

	2010	2012			2014		
	Nonparametric(1)		2SLS(2)		NP(1)	2SLS(2)	
Bandwidth	Varies	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%
Teacher Improvement							
Professional development	0.259	0.537+	0.385***	0.311*	0.546*	0.486*c	0.101
	(0.203)	(0.280)	(0.114)	(0.139)	(0.260)	(0.206)	(0.128)
Individual planning	0.152	-0.121	0.045c	-0.238	0.296	-0.169	-0.144
	(0.388)	(0.352)	(0.368)	(0.188)	(0.372)	(0.174)	(0.211)
Collaborative planning	1.263***	0.579*	0.186	0.163	1.031***	0.023q	0.045
	(0.334)	(0.261)	(0.115)	(0.148)	(0.311)	(0.164)	(0.129)
Utilizing results of assessments	0.377	0.609*	-0.096	-0.163	-0.077	-0.096	0.052
	(0.449)	(0.237)	(0.115)	(0.154)	(0.256)	(0.115)	(0.145)
Administrative Burdens							
Supervisory duties	-0.105	0.332	0.421*q	0.270+	0.165	0.073	0.122
	(0.275)	(0.326)	(0.191)	(0.155)	(0.214)	(0.106)	(0.125)
Required committee/staff meetings	0.198	0.140	0.369**	0.288+	0.761***	0.343**	0.257+
	(0.220)	(0.261)	(0.125)	(0.156)	(0.231)	(0.117)	(0.151)
Completing required paperwork	0.209	0.480+	0.309*q	0.224+	0.247	0.001	0.476*q
	(0.220)	(0.286)	(0.167)	(0.130)	(0.214)	(0.106)	(0.187)
Community & Students							
Communicating with parents/community	0.364**	0.333+	-0.038q	-0.079	0.537*	0.100	0.333+q
	(0.137)	(0.193)	(0.109)	(0.091)	(0.220)	(0.085)	(0.180)
Addressing student discipline	0.091	0.320	0.099	0.304q	0.724	0.282	0.675q
	(0.252)	(0.355)	(0.164)	(0.337)	(0.474)	(0.188)	(0.413)
Focusing on Tests							
Prep for federal, state, and local tests	0.336	0.870*	0.036	0.121	0.439*	0.053	0.139
	(0.269)	(0.365)	(0.141)	(0.181)	(0.214)	(0.145)	(0.173)
Delivery of assessments	0.174	0.770***	-0.028	-0.011	0.422*	0.193+	0.606*q
	(0.251)	(0.223)	(0.099)	(0.138)	(0.189)	(0.115)	(0.255)
N	1753	1753	518	294	1753	518	294
Controls for 2010 baseline composite?	YES	YES	YES	YES	YES	YES	YES
Controls for 2010 outcome & school level?	NO	NO	YES	YES	NO	YES	YES
Includes baseline observations?	NO	NO	NO	NO	NO	NO	NO

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Nonparametrics bandwidths calculated from Imbens and Kalyanaraman (2011).

(2) Linear spline equation used in parametric 2SLS models unless otherwise noted; *q* =quadratic equation used; *c*= cubic equation used.

Table 6: School Climate as Perceived by Teachers; Estimates by Method, Bandwidth, and Year

		Math(1)					Reading(1)				
Subgroup (based on 2010 Score):		All	Level I	Level II	Level III	Level IV	All	Level I	Level II	Level III	Level IV
2012 Passing Rates											
+/- 16%(2)		0.352	11.695	11.273	0.285	-1.506	-3.314	1.500	-3.612	-1.164	-3.184
		(2.498)	(22.177)	(7.857)	(2.539)	(1.949)	(3.188)	(5.093)	(7.711)	(3.455)	(3.240)
	N	23862	1355	5614	13667	3226	23865	6520	5419	9651	2275
+/- 10%(2)		4.879+	1.067	21.034*	4.459	-1.640	-0.919	4.574	-1.857	1.227	-4.739
		(2.786)	(25.097)	(9.570)	(2.790)	(2.025)	(3.838)	(5.985)	(9.213)	(4.154)	(4.622)
	N	13190	890	3482	7410	1408	13194	4079	3086	5017	1012
+/- 5%(2)		-0.508	-11.317	17.323	-1.028	-1.437	-7.198	9.598	-13.396	-5.260	-9.205
		(4.283)	(39.718)	(13.447)	(4.527)	(1.457)	(6.402)	(10.009)	(14.206)	(6.946)	(7.794)
	N	5766	397	1637	3166	566	5770	1866	1374	2131	399
2012 Standardized Scores											
+/- 16%(2)		0.005	-0.423	0.155	0.008	0.025	-0.016	0.047	0.015	0.001	-0.393*
		(0.069)	(0.373)	(0.109)	(0.071)	(0.147)	(0.049)	(0.099)	(0.086)	(0.057)	(0.175)
	N	23398	1143	5410	13620	3225	23277	5988	5369	9645	2275
+/- 10%(2)		0.086	-0.397	0.289*	0.083	0.121	0.025	0.177	0.083	0.017	-0.356+
		(0.076)	(0.430)	(0.135)	(0.076)	(0.166)	(0.061)	(0.124)	(0.103)	(0.070)	(0.190)
	N	12887	755	3348	7377	1407	12822	3737	3057	5016	1012
+/- 5%(2)		-0.035	0.650	0.127	-0.052	0.179	-0.130	0.305	-0.172	-0.157	-0.641*
		(0.125)	(0.696)	(0.175)	(0.130)	(0.246)	(0.105)	(0.205)	(0.177)	(0.119)	(0.273)
	N	5639	346	1576	3152	565	5610	1720	1361	2130	399

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Columns split into all students from 2010 and separate analyses by 2010 category. Level I and II represent failing ratings. N lower for test scores than passing rates; small number of missing test scores retained score category.

(2) Analysis uses linear 2SLS models for students who were in treated and untreated schools within the given cutoff in the baseline year. All models control for the school-level baseline composite score, student-level baseline math scores, student-level baseline reading scores, and interactions between these continuous variables, an indicator for being below the assignment score (creating a spline), and the baseline outcome level (to allow for different relationships in the data for different levels of ability). The analysis clusters standard errors for the student's 2010 school. If anything, results are stronger without controlling for both tests; we include both tests to be conservative.

Table 7: School-level Student Composition; Estimates by Method and Year

	2012			2013			2014		
	NP(1)	2SLS(2)		NP(1)	2SLS(2)		NP(1)	2SLS(2)	
Bandwidth	Varies	+/-16%	+/-8%	Varies	+/-16%	+/-8%	Varies	+/-16%	+/-8%
Percent FRL Students	4.652+	2.842*	3.886*	5.020+	2.415	3.881*	5.996*	3.427*	4.197*
	(2.654)	(1.447)	(1.748)	(2.999)	(1.484)	(1.754)	(2.938)	(1.515)	(1.731)
Percent Black Students	5.227	0.596	-0.004	7.719	0.596	1.880	9.377	1.881	2.135
	(5.216)	(0.966)	(1.259)	(6.942)	(0.966)	(1.522)	(7.436)	(1.335)	(1.717)
Percent Hispanic Students	-2.734	-0.276q	-0.032	-3.429	-0.180	-0.428	-4.220	-0.529	-1.295
	(3.985)	(1.138)	(1.026)	(3.747)	(0.948)	(1.194)	(4.084)	(1.013)	(1.225)
N	1753	518	294	1753	518	294	1753	518	294
Controls for 2010 baseline composite?	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls for 2010 outcome & school level?	NO	YES	YES	NO	YES	YES	NO	YES	YES
Includes baseline observations?	NO	NO	NO	NO	NO	NO	NO	NO	NO

+ p-value<0.1, * p-value<0.05, ** p-value<0.01, *** p-value<0.001

(1) Nonparametric bandwidths calculated from Imbens and Kalyanaraman (2011).

(2) Linear spline equation used in parametrics models unless otherwise noted; *q* =quadratic equation used; *c*= cubic equation used.

Table A1: Survey Items and Factors

Construct	Question	RD 2012 +/-16%	RD 2012 +/-8%	RD 2014 +/-16%	RD 2014 +/-8%
School Leadership	Teachers are recognized as educational experts.	-0.069 (0.063)	-0.038 (0.092)	-0.059 (0.070)	-0.087 (0.084)
	Teachers are trusted to make sound professional decisions about instruction.	-0.074 (0.069)	-0.039 (0.099)	-0.051 (0.082)	-0.029 (0.098)
	Teachers are relied upon to make decisions about educational issues.	-0.066 (0.063)	-0.025 (0.088)	-0.034 (0.072)	-0.035 (0.090)
	Teachers are encouraged to participate in school leadership roles.	-0.014 (0.052)	-0.005 (0.076)	0.004 (0.054)	-0.013 (0.062)
	The faculty has an effective process for making group decisions to solve problems.	-0.004 (0.072)	0.028 (0.106)	-0.013 (0.076)	-0.033 (0.089)
	In this school we take steps to solve problems.	-0.019 (0.069)	-0.006 (0.100)	-0.018 (0.080)	-0.031 (0.097)
	Teachers are effective leaders in this school.	-0.040 (0.055)	-0.009 (0.081)	-0.009 (0.065)	-0.019 (0.075)
	Teachers have an appropriate level of influence on decision making in this school.	-0.063 (0.067)	0.013 (0.098)	-0.026 (0.074)	-0.069 (0.084)
	The faculty and staff have a shared vision.	-0.021 (0.067)	0.034 (0.099)	-0.012 (0.075)	-0.074 (0.086)
	There is an atmosphere of trust and mutual respect in this school.	-0.053 (0.088)	0.032 (0.130)	-0.021 (0.095)	-0.064 (0.110)
	Teachers feel comfortable raising issues and concerns that are important to them.	-0.023 (0.090)	0.069 (0.131)	0.063 (0.093)	0.036 (0.111)
	The school leadership consistently supports teachers.	-0.044 (0.084)	0.018 (0.121)	0.024 (0.091)	-0.011 (0.107)
	Teachers are held to high professional standards for delivering instruction.	-0.015 (0.045)	-0.048 (0.063)	-0.051 (0.056)	-0.093 (0.068)
	Teacher performance is assessed objectively.	-0.039 (0.063)	-0.045 (0.088)	0.006 (0.070)	-0.013 (0.085)
	Teachers receive feedback that can help them improve teaching.	-0.041 (0.067)	-0.074 (0.095)	-0.008 (0.078)	-0.082 (0.096)
	The procedures for teacher evaluation are consistent.	-0.058 (0.073)	-0.064 (0.099)	-0.068 (0.086)	-0.112 (0.096)
	The school improvement team provides effective leadership at this school.	-0.065 (0.067)	-0.046 (0.100)	-0.027 (0.068)	-0.069 (0.081)
	The faculty are recognized for accomplishments.	-0.013 (0.073)	0.008 (0.106)	0.029 (0.071)	-0.047 (0.085)
	The school leadership makes a sustained effort to address teacher concerns about: Leadership issues	-0.033 (0.070)	0.018 (0.102)	-0.033 (0.073)	-0.056 (0.088)
	The school leadership makes a sustained effort to address teacher concerns about: Facilities and resources	-0.042 (0.059)	-0.031 (0.084)	-0.044 (0.067)	-0.104 (0.080)
	The school leadership makes a sustained effort to address teacher concerns about: The use of time in my school	-0.057 (0.069)	-0.024 (0.100)	-0.009 (0.074)	-0.050 (0.089)
	The school leadership makes a sustained effort to address teacher concerns about: Professional development	-0.105+ (0.064)	-0.083 (0.091)	-0.066 (0.065)	-0.114 (0.079)

	The school leadership makes a sustained effort to address teacher concerns about: Teacher leadership	-0.025 (0.061)	-0.006 (0.089)	-0.054 (0.067)	-0.098 (0.082)
	The school leadership makes a sustained effort to address teacher concerns about: Community support and involvement	-0.043 (0.058)	-0.020 (0.083)	-0.006 (0.067)	-0.039 (0.079)
	The school leadership makes a sustained effort to address teacher concerns about: Managing student conduct	-0.029 (0.077)	0.038 (0.112)	0.008 (0.080)	-0.021 (0.095)
	The school leadership makes a sustained effort to address teacher concerns about: Instructional practices and support	-0.049 (0.060)	-0.047 (0.085)	-0.040 (0.066)	-0.097 (0.081)
	The school leadership makes a sustained effort to address teacher concerns about: New teacher support	-0.039 (0.071)	0.035 (0.100)	-0.035 (0.076)	-0.024 (0.100)
	Teachers are encouraged to try new things to improve instruction.	-0.004 (0.047)	-0.004 (0.066)	-0.009 (0.047)	-0.024 (0.058)
	Teachers are assigned classes that maximize their likelihood of success with students.	-0.011 (0.070)	-0.020 (0.098)	-0.020 (0.066)	-0.003 (0.082)
	Teachers have autonomy to make decisions about instructional delivery (i.e. pacing, materials and pedagogy).	-0.082 (0.065)	-0.061 (0.090)	0.011 (0.062)	-0.004 (0.082)
	Overall, my school is a good place to work and learn.	-0.061 (0.066)	-0.043 (0.102)	-0.041 (0.084)	-0.082 (0.100)
Instructional Practices	The school leadership facilitates using data to improve student learning.	-0.034 (0.050)	-0.062 (0.071)	-0.051 (0.057)	-0.077 (0.073)
	State assessment data are available in time to impact instructional practices.	0.030 (0.045)	-0.009 (0.061)	-0.078 (0.057)	-0.066 (0.078)
	Local assessment data are available in time to impact instructional practices.	0.021 (0.045)	0.011 (0.065)	-0.078 (0.051)	-0.106+ (0.063)
	Teachers use assessment data to inform their instruction.	0.003 (0.034)	-0.002 (0.050)	-0.056 (0.039)	-0.094+ (0.049)
	Teachers work in professional learning communities to develop and align instructional practices.	-0.017 (0.047)	-0.043 (0.066)	-0.044 (0.050)	-0.083 (0.061)
	Provided supports (i.e. instructional coaching, professional learning communities, etc.) translate to improvements in instructional practices by teachers.	-0.013 (0.047)	0.013 (0.063)	-0.014 (0.050)	-0.064 (0.059)
Professional Development	Sufficient resources are available for professional development in my school.	-0.030 (0.055)	-0.037 (0.072)	-0.035 (0.065)	-0.091 (0.074)
	An appropriate amount of time is provided for professional development.	-0.009 (0.053)	-0.048 (0.070)	-0.001 (0.058)	-0.051 (0.068)
	Professional development offerings are data driven.	0.014 (0.056)	-0.000 (0.078)	-0.012 (0.049)	-0.032 (0.062)
	Professional learning opportunities are aligned with the school’s improvement plan.	-0.041 (0.047)	-0.030 (0.064)	-0.007 (0.052)	-0.081 (0.061)
	Professional development is differentiated to meet the individual needs of teachers.	-0.063 (0.065)	0.008 (0.088)	-0.054 (0.070)	-0.137 (0.084)
	Professional development deepens teachers' content knowledge.	-0.042 (0.055)	-0.047 (0.074)	-0.038 (0.055)	-0.104 (0.069)
	Teachers have sufficient training to fully utilize instructional technology.	-0.107+ (0.063)	-0.051 (0.087)	-0.020 (0.060)	-0.065 (0.073)
	Teachers are encouraged to reflect on their own practice.	-0.015 (0.042)	-0.007 (0.056)	0.001 (0.045)	-0.036 (0.053)

	In this school, follow up is provided from professional development.	-0.032 (0.062)	-0.009 (0.086)	-0.059 (0.070)	-0.149+ (0.081)
	Professional development provides ongoing opportunities for teachers to work with colleagues to refine teaching practices.	-0.037 (0.057)	-0.027 (0.078)	-0.041 (0.058)	-0.105 (0.070)
	Professional development is evaluated and results are communicated to teachers.	-0.053 (0.067)	-0.052 (0.093)	-0.027 (0.064)	-0.073 (0.084)
	Professional development enhances teachers' ability to implement instructional strategies that meet diverse student learning needs.	-0.040 (0.053)	-0.024 (0.073)	-0.042 (0.054)	-0.086 (0.070)
	Professional development enhances teachers' abilities to improve student learning.	-0.054 (0.050)	-0.040 (0.068)	-0.047 (0.052)	-0.125+ (0.066)
Community-School Relations	Parents/guardians are influential decision makers in this school.	-0.060 (0.059)	-0.026 (0.084)	-0.066 (0.078)	-0.137 (0.100)
	This school maintains clear, two-way communication with the community.	-0.035 (0.055)	0.004 (0.082)	-0.009 (0.065)	-0.052 (0.087)
	This school does a good job of encouraging parent/guardian involvement.	-0.045 (0.058)	-0.026 (0.084)	-0.019 (0.069)	-0.068 (0.092)
	Teachers provide parents/guardians with useful information about student learning.	-0.015 (0.035)	-0.011 (0.047)	-0.029 (0.044)	-0.054 (0.057)
	Parents/guardians know what is going on in this school.	-0.018 (0.054)	-0.005 (0.081)	-0.015 (0.062)	-0.015 (0.080)
	Parents/guardians support teachers, contributing to their success with students.	0.032 (0.053)	0.029 (0.077)	0.007 (0.065)	-0.047 (0.085)
	Community members support teachers, contributing to their success with students.	0.051 (0.055)	0.061 (0.082)	-0.085 (0.073)	-0.116 (0.097)
	The community we serve is supportive of this school.	-0.015 (0.063)	-0.032 (0.092)	-0.051 (0.070)	-0.098 (0.096)
Student Conduct	Students at this school understand expectations for their conduct.	-0.039 (0.066)	-0.027 (0.092)	0.014 (0.076)	0.002 (0.088)
	Students at this school follow rules of conduct.	-0.055 (0.087)	-0.071 (0.121)	-0.046 (0.102)	-0.060 (0.126)
	Policies and procedures about student conduct are clearly understood by the faculty.	0.002 (0.060)	0.053 (0.085)	-0.037 (0.071)	-0.040 (0.085)
	School administrators consistently enforce rules for student conduct.	0.015 (0.099)	0.113 (0.138)	-0.023 (0.108)	-0.011 (0.131)
	School administrators support teachers' efforts to maintain discipline in the classroom.	0.016 (0.095)	0.077 (0.131)	0.000 (0.094)	0.010 (0.114)
	Teachers consistently enforce rules for student conduct.	0.003 (0.045)	0.012 (0.062)	-0.063 (0.051)	-0.096 (0.062)
	The faculty work in a school environment that is safe.	-0.038 (0.059)	0.007 (0.085)	-0.073 (0.070)	-0.067 (0.082)
Facilities & Resources	Teachers have sufficient access to appropriate instructional materials.	-0.108 (0.069)	-0.144+ (0.085)	-0.042 (0.074)	-0.075 (0.096)
	Teachers have sufficient access to instructional technology, including computers, printers, software and internet access.	-0.113 (0.089)	-0.118 (0.125)	-0.019 (0.079)	-0.017 (0.102)
	Teachers have access to reliable communication technology, including phones, faxes and email.	-0.093 (0.062)	-0.095 (0.085)	-0.068 (0.064)	-0.090 (0.079)

	Teachers have sufficient access to office equipment and supplies such as copy machines, paper, pens, etc.	-0.056 (0.076)	-0.134 (0.102)	-0.050 (0.087)	-0.169 (0.105)
	Teachers have sufficient access to a broad range of professional support personnel.	-0.041 (0.054)	-0.011 (0.075)	0.004 (0.057)	-0.017 (0.069)
	The school environment is clean and well maintained.	-0.116+ (0.066)	-0.096 (0.097)	-0.059 (0.075)	-0.043 (0.101)
	Teachers have adequate space to work productively.	-0.090+ (0.055)	-0.150* (0.072)	-0.080 (0.051)	-0.113 (0.069)
	The physical environment of classrooms in this school supports teaching and learning.	-0.106* (0.054)	-0.126+ (0.076)	-0.060 (0.055)	-0.108 (0.071)
	The reliability and speed of Internet connections in this school are sufficient to support instructional practices.	-0.074 (0.074)	-0.098 (0.104)	-0.099 (0.078)	-0.133 (0.099)
Time	Class sizes are reasonable such that teachers have the time available to meet the needs of all students.	-0.100 (0.087)	-0.100 (0.087)	-0.075 (0.091)	-0.075 (0.091)
	Teachers have time available to collaborate with colleagues.	-0.090 (0.071)	-0.133 (0.096)	-0.109 (0.068)	-0.166+ (0.090)
	Teachers are allowed to focus on educating students with minimal interruptions.	-0.065 (0.072)	-0.090 (0.098)	-0.154+ (0.085)	-0.196+ (0.102)
	The non-instructional time provided for teachers in my school is sufficient.	-0.108 (0.077)	-0.110 (0.112)	-0.138 (0.087)	-0.182+ (0.107)
	Efforts are made to minimize the amount of routine paperwork teachers are required to do.	-0.087 (0.071)	-0.163+ (0.097)	-0.103 (0.077)	-0.247** (0.093)
	Teachers have sufficient instructional time to meet the needs of all students.	-0.040 (0.053)	-0.133+ (0.072)	-0.145* (0.062)	-0.196* (0.082)
	Teachers are protected from duties that interfere with their essential role of educating students.	-0.141* (0.066)	-0.204* (0.087)	-0.116+ (0.069)	-0.206* (0.084)

[1] Additionally, the state must: (1) ensure that all TALAS schools and districts receive school- and district-specific support to increase student achievement, (2) require districts to focus on the lowest-achieving schools, (3) increase strategies and options in TALAS plans, and (4) develop several STEM high school networks (RttT Application, 2010). Steps 1-3 apply to all TALAS schools, while Step 4 pertains to high schools.

[2] For instance, one school implemented a 1:1 laptop initiative, a K-5 STEM program, and digital literacy programs, while another implemented weekly meetings for Algebra I teachers to plan lessons and a focus on individualized literacy improvement plans for students 3 grades below level (Department of Public Instruction, 2013a).

[3] Ninety percent of the Regional Leadership Academy graduates were placed in a "high-needs" school by October 2013 (Department of Public Instruction, 2013b), though it's not clear that these were necessarily turnaround schools. Some professional development materials for school leaders are available here: http://dst.ncdpi.wikispaces.net/PD+for+School+Leaders .

[4] The state sends a link to an online survey to every educator in the state in the spring of every evenly-numbered year. The mean response rates were 90.3% in 2010, 88.5% in 2012, and 92.2% in 2014. Controlling for response rates does not change our results. All schools had at least one response in 2010 and 2012, while one treatment and one control school were missing responses in 2014 (0.4% of the main data we examine). We replace the missing 2014 data with the 2012 value in our main analysis; dropping the missing schools does not change our results.

[5] We identify a change in school principal by using the NCERDC data on educator-level pay. When schools had more than one principal in a given year, we treated the principal with the most months in the school in that year as the principal of record. If multiple principals had equal time, we took the principal who started the year as the principal of record. If the school was missing a principal in a given year, we assumed the principal from the prior year remained in the school (that is, we assumed no turnover). In 2010, a quirk in the data led to 96 schools, or 5.4% of the total schools, missing teacher turnover data. We used the 2009 estimate as the baseline teacher turnover for 62 of the schools; the remaining 34 schools had just opened in 2010 and thus had no turnover relative to 2009. No schools were missing other school-level DPI data in any year.

[6] There were 66 treated elementary schools (5% of 1,321) and 23 treated middle schools (5% of 451).

[7] Alternatively, perhaps NCDPI manipulated the threshold in order to usher particular schools into the program. The 5% cutoff is a federal standard, and the state would have little room for shifting schools. Though it seems unlikely, we cannot rule out this possibility. Importantly, such manipulation would constitute an internal validity problem only if NCDPI selected schools that had similar outcomes on the assignment variable but for some unobserved reason had a higher likelihood of positive (or negative) outcomes under the treatment (Dee, 2012).

[8] We ran various tests to check for manipulation. First, if no manipulation occurred the distribution of schools by composite score should have a normal distribution. Using methods suggested by McCrary (2008), we examine whether there is a break in the distribution at the cutoff. The small difference is not statistically significant at traditional levels of confidence (coefficient=6.2 schools, p-value=0.193), indicating that there is no jump in density.

[9] In some specifications, the parametric RD models include the baseline level of the outcome variable and school type. Including this control has no effect on the overall results but increases the precision of the estimates.

[10] Lee and Lemieux (2010) suggest starting with a linear model, inserting bin indicator variables into the polynomial regression, and jointly testing their significance. For instance, we placed K-2 bin indicators (each two percentage points wide), B_k, for $k = 2$ to $K - 1$, into our model above:

$$(3) \qquad Y_s = \pi\, \widehat{Turnaround}_s + g(A_s) + \beta X_s + \sum_{k=2}^{K-1} \varphi_k B_k + \varepsilon_s$$

We then tested the null hypothesis that $\varphi_2 = \varphi_3 = \ldots = \varphi_{K-1} = 0$. Starting with a first order polynomial (flexible across the discontinuity), we added a higher order to the model until the bin indicator variables were no longer jointly significant. This method also tests for discontinuities at unexpected points along the assignment variable; we did not find any. We limit the flexibility to a third-order polynomial.

[11] Using a pure intent-to-treat analysis (i.e., asking, given that your 2010 school was below the cutoff, how did you do relative to a student whose 2010 school was above the cutoff?) gives functionally the same results. Alternatively, we could predict the student-level probability of attending a treated school in fifth grade in 2012 for a treatment-on-the-treated analysis.

[12] In theory, we could examine whether the treatment effect is constant below the cut point by examining whether the treated and untreated dashed lines are parallel (Tang, Cook, & Kisbu-Sakarya, 2014; Wing & Cook, 2013). Indeed, it appears that the drop in scores was smaller at very bottom-scoring schools. However, we are apprehensive

about making generalizations beyond the cut point in our context, both because the lowest-achieving schools had less distance to fall and because other programs may have affected schools away from out cut point.

[13] The first stage coefficient may change slightly from estimate-to-estimate, as the IK bandwidths change in nonparametric estimates and the baseline controls differ depending on the outcome variable in parametric estimates. The first stage displayed is for the first listed outcome variable in the table.

[14] Not all treatment schools replaced their principal. Schools were exempted from the replacement requirement if they had recently replaced their principal as part of the earlier turnaround program and the school had made substantial improvements on their composite score during the new principal's tenure (Henry et al., 2014).

[15] We use the state-defined variable for teacher turnover, which is the number of teachers who were employed in March of the previous year (Year 0) but who were not employed the following year (Year 1), divided by the total teachers who were employed in March of the previous year (Year 0).

[16] Schools differed substantially in what they included in their annual reports, and many schools who mentioned teacher action plans in 2012 did not mention the results of those plans in 2013. Other schools did not mention teacher action plans in 2012, but do note that they began the process of replacing teachers for low performance in 2013. In 2013, one school notes that "five teachers whose performance concerned the principal resign mid-year. Four of those teachers were not hired by the principal but were assigned to the school by the central office."

[17] Certain schools mention programs like Child Family Support Teams comprised of the school nurse, guidance counselor, social worker, and administrators that attempt to connect families to community resources (Department of Public Instruction, 2013a). Other schools use backpack programs to provide food over the weekend for low-income children. However, because schools design their own programs, these are not present in every school, and some of these programs may have existed even before TALAS. Future research should systematically review these programs to understand what effect, if any, they may have.