School Turnaround in North Carolina:

A Regression Discontinuity Analysis

Jennifer A. Heissel[a*] and Helen F. Ladd[b]

[a]Graduate School of Business and Public Policy
555 Dyer Rd.
Naval Postgraduate School
Monterey, CA 93943

[b]Sanford School of Public Policy
201 Science Dr.
Duke University
Durham, North Carolina 27708

*Corresponding author: jaheisse@nps.edu

Acknowledgements.: Helen Ladd is grateful to IES CALDER grant R305C12008 for financial assistance.

Jennifer A. Heissel, Helen F. Ladd. School turnaround in North Carolina: A regression discontinuity analysis. Economics of Education Review, Volume 62, 2018, Pages 302-320.

School Turnaround in North Carolina: A Regression Discontinuity Analysis

Abstract

This paper examines the effect of a federally supported school turnaround program in North Carolina elementary and middle schools. Using a regression discontinuity design, we find that the turnaround program did not improve, and may have reduced, average school-level passing rates in math and reading. One potential contributor to that finding appears to be that the program increased the concentration of low-income students in treated schools. Based on teacher survey data, we find that, as was intended, treated schools brought in new principals and increased the time teachers devoted to professional development. At the same time, the program increased administrative burdens and distracted teachers, potentially reducing time available for instruction, and increased teacher turnover after the first full year of implementation. Overall, we find little evidence of success for North Carolina's efforts to turn around low-performing schools under its Race to the Top grant.

Keywords

Accountability; educational reform; elementary schools; middle schools; regression discontinuity; state and federal aid

Highlights

- Paper examines school turnaround in North Carolina elementary and middle schools.
- Turnaround efforts did not improve and may have reduced math and reading scores.
- Program increased administrative burdens.

JEL Classification

I21

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 and student behavior problems to poor school leadership and 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 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 the problems of these schools. Instead what is needed, according to this view, are broader whole-school reform efforts that comprehensively address the range of problems such schools face, including 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, 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 on schools that are close to the cut off. 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 (North Carolina Race to the Top Application, 2010). We find, however, that the turnaround program did not increase average achievement at either the school or the student level during the first few years after the program was implemented. Instead we find that passing rates at best stayed the same and may have fallen. For reasons we discuss below, this negative finding differs from more positive findings that emerged from previous research on this program (Henry, Campbell, Thompson, & Townsend, 2014; Henry, Guthrie, & Townsend, 2015).

Although we cannot pinpoint the specific causes of the disappointing student outcomes, we were able to explore a number of both intended and unintended consequences of the turnaround strategy that could have contributed to them. We find, for example, no evidence that the turnover of principals, which was a central part of the strategy, increased the quality of school leadership in the schools subject to turnaround. Consistent with the intent of the program, we find that teachers devoted more time to professional development, but that they also faced more administrative burdens, with no perceived improvement in school climate. An unintended outcome was an increase in the share of low-income students in the turnaround schools.

2. Background and prior policy research

Individual states, including North Carolina, have long used a variety of approaches to turn around their lowest performing schools. Their efforts have been bolstered in recent years by $7 billion dollars of federal funding in the form of Race to the Top (RttT) and School Improvement Grants (Dragoset et al., 2016, 2017). States that received federal grants to improve their lowest achievement schools were required to employ one of the following four specific models:

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.

Both nationwide and in North Carolina, the majority of schools that received funding selected either the transformation or the turnaround model. Central to these preferred models are the replacement of the principal and the improvement of teachers.

Concern about the quality of school leadership reflects the central role that principals play in schools as they make personnel decisions, set policies and practices, distribute leadership authority, and influence school culture. Research documents that principals vary in their effectiveness, especially in high-poverty schools (Branch, Hanushek, & Rivkin, 2012). By calling for the replacement of principals in low-performing schools, federal policymakers expected the new principals to be more successful than the ones they replaced. However, replacing an experienced principal with an inexperienced one may bring few benefits and could be counterproductive (Clark, Martorell, & Rockoff, 2009).

With a new principal a school 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). Evidence shows that this approach can develop the school-wide capacity-building and ownership needed to sustain school reforms (Copland, 2003)

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 statistically control 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, the turnaround model requires a school to replace 50% of its teachers. The usefulness of this policy depends in part on the quality of the replacement teachers. Such a requirement, for example, may pose a challenge for rural areas with a limited supply of qualified teachers to replace those who are fired (Cowen, Butler, Fowles, Streams, & Toma, 2012; Sipple & Brent, 2007). On a more positive note, some research has shown that changing the group of teachers in a school can improve their joint productivity in low-performing schools (Hansen, 2013).

The transformation and turnaround models also call for more investment in the professional development of teachers, a strategy that can be productive provided the program is high-quality (Hill, 2007). However, 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 evidence that principal and teacher quality matter, whether comprehensive school turnaround strategies of the type promoted by the federal government will improve the lowest achieving schools is an empirical question. A review by the What Works Clearinghouse, 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 all low-performing schools. A recent national study by the U.S. Department of Education found that the School Improvement Grants generated no benefits to student outcomes (Dragoset et al., 2017).

More positive evidence emerges from a set of studies of the same North Carolina program that we investigate in this paper (Henry et al., 2014, 2015). In contrast to the regression discontinuity approach that we use, these prior North Carolina studies rely on a difference-in-difference (DID) approach. In a concluding discussion, we reconcile our far less positive results with the positive findings from these earlier studies and argue that our RD approach is the preferred approach for estimating the causal impacts of the program.

3. North Carolina Data and Policy Context

North Carolina has been engaged in school turnaround efforts for over 10 years, with much of its attention focused on low performing high schools.[1] Drawing on that experience, the state successfully competed for federal Race to the Top (RttT) 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.

We use data for elementary and middle schools in the 2010 through 2014 school years from the North Carolina Department of Public Instruction (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.[2] We separately analyze the time use and school climate measures from the survey. 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

[1] Created in 2006, the District and School Transformation department 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).

[2] 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%. 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.

conduct, school facilities and resources, and time use. This method results in a Z-score (with an average of zero and a standard deviation of one) for each factor in each school by year. See Appendix A 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.[3] 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.

NCDPI based assignment to treatment on a school's 2010 composite score, calculated as the number of passing scores on reading, mathematics, science, and end-of-course tests as a percent of all such tests taken in the school.[4] 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. We limit our analysis to elementary and middle schools, in part because their cut point for assignment to the program was based on test scores alone and was not complicated by the inclusion of graduation rates. Leaving out high schools also reduces the potential for confounding the effects of TALAS with the more intensive high school intervention from the previous state-sponsored program. We exclude

[3] 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 NCDPI data in any year.

[4] Calculated by authors using NCDPI rules.

private, charter, alternative, and special education schools, because they were not eligible for TALAS.

Eighty-nine elementary and middle schools out of a total 1,772 public elementary and secondary schools met the eligibility criterion in 2010.[5] Four treated 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. We require schools to appear in all years 2010-2014 to be included in the analysis. Including schools before they closed did not change the results.

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).[6] By the end of the 2011 school year, all TALAS schools had taken some steps of an intervention model, but many of these efforts had not yet been fully implemented (Whalen, 2011). 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. No schools chose the restart model.

In summer 2011, 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

[5] There were 66 treated elementary schools (5% of 1,321) and 23 treated middle schools (5% of 451).
[6] Additionally, the state had to: (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 (North Carolina Race to the Top Application, 2010). Steps 1-3 applied to all TALAS schools, while Step 4 pertained to high schools.

(Department of Public Instruction, 2013b; Henry et al., 2014, 2015). Coaches generally served more than one school, with an average of about one day per week spent at a given TALAS school (Henry et al., 2014). The particular strategies employed by the coaches differed by school.[7] In general the leadership coaching strategies used 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 discussed the professional development provided to principals and teachers, with a particular emphasis on school and teacher leadership, as well as teacher recruitment efforts by principals (Department of Public Instruction, 2013b, 2014).[8] Schools continued these strategies in the 2013 and 2014 school years. Our analysis follows schools, students, and teachers through 2014.

The school-level TALAS program took place in the context of additional RttT-funded reforms in North Carolina, including a district-level turnaround program run by the state's District and School Transformation department that had been established in 2007. This group viewed the district as an important unit for change because districts make important policy and personnel decisions, including principal staffing decisions. We focus here on the school-level TALAS program, but schools above and below the school-level cut point also could have received this district-level support.

4. 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

[7] 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).

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

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 differ slightly to ensure that 5% of each school type is included in TALAS. By centering each school's composite score on 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 in two-percentage point bins.

The main takeaway from Figure 1 is the strong discontinuity in uptake at the cut point. We note, however, that two schools above the cut point did not comply with their assignments. 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, which suggests that 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). The intended treatment population includes those below the cutoff and the intended control population includes those above that point. This 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. The estimates should be interpreted as a *local average treatment effect* (LATE, Angrist, Imbens, & Rubin, 1996; Angrist & Pischke, 2009; Hahn, Todd, & Van der Klaauw, 2001). In other words, the estimate is only for those whose uptake is affected by the assignment around the cut point.

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.

Second, specifying the correct functional form presents a challenge. We present a variety of specifications for each outcome of interest, using both nonparametric and parametric methods (Lee & Lemieux, 2010). The nonparametric estimates are 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. For the parametric analysis, we implement a fuzzy RD design with a two-stage parametric model that functions as an instrumental variable analysis (Hahn et al., 2001; Lee & Lemieux, 2010; Van Der Klaauw, 2008).[9] Because we do not know the "true" relationship between the outcome and the assignment variable, we cannot be certain whether the functional form should be linear, quadratic, cubic, or something else entirely. We follow a test proposed by Lee and Lemieux (2010) to find the best-fitting parametric form.[10] The models that follow use the simplest form not rejected by this test; the vast majority have a linear spline on either side of the cutoff. Appendix B includes additional details on the specifications.

---

[9] The first-stage model estimates the jump in treatment probability at the cutoff point with the following form:

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

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 to reduce sampling variability. 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) $$Y_s = \pi\, \widehat{Turnaround}_s + g(A_s) + \beta X_s + \varepsilon_s$$

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 X_s included in Model 1. Under assumptions of monotonicity and excludability, this system of equations functions as an instrumental variable estimate (Angrist, Imbens, & Rubin, 1996; Angrist & Pischke, 2009; Hahn, Todd, & Van der Klaauw, 2001).

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

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.

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 parametric models with varying bandwidths. We use +/-16 percentage points from the composite score cut point 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. We also report results based on a bandwidth of +/-10 percentage points.

The 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).[11] In any case, we find no empirical evidence of such manipulation.[12]

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

[11] Alternatively, perhaps NCDPI manipulated the threshold 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).

[12] 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.

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). Panel B uses the same parametric function described above. 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 documented relationship between student test scores and various measures of disadvantage. Schools below the cut point are also more likely to have been in the 2007 DST school turnaround program and be assigned to the 2011 RttT district-level program. These differences indicate that a simple comparison of schools above and below the cutoff, as in the Henry et al. (2014, 2015) papers, could produce biased estimates of the effects of the policy intervention. However, when we focus on a comparison of schools at the cutoff point (as in Panel B), the differences disappear. See Appendix C for additional details about differences in programs away from the cut point.

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.

5. Student Outcomes

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 but follows them for only two years because after that they move to middle schools that may or may not be treated.

Figure 2 displays the composite, math, and reading outcomes based on a simple model with no additional control variables. The gray line is the 2010 baseline trend, the solid black line is the 2014 segment for schools intended as controls, and the dashed black line is the 2014 segment for schools intended for treatment.[13] The significant decline in passing rates between 2010 and 2014 (see the difference between the gray and the black lines) reflects the fact that the state changed its tests and raised the corresponding passing standards during the period. This change in standards, however, should not interfere with our estimates of the program effects, which are measured at the 2010 cut point (denoted by zero in the figure) for schools that are virtually identical in all measurable dimensions. Contrary to expectations, the figure indicates that at the cut point, the 2014 passing rates are lower in the treated than in the control schools.

More formally, but generally consistent with the figure, Table 2 provides no evidence that the program had positive effects on school wide pass rates overall or for various subgroups defined by gender, race, or FRL status. Results are reported by post-program year and for various model specifications. The first row of the table provides the first stage estimate of the increase in assignment to the treatment caused by the discontinuity.[14] As expected, there is a strong uptick in

[13] 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, 2015; Wing & Cook, 2013). Indeed, it appears that the drop in scores was smaller at very low 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 (see Appendix C).

[14] 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 overall math estimate.

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 pass rates are in the following rows. We highlight outcomes that are significant at p<0.10 for a majority of estimates. 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 2013 and 2014. Our preferred estimates are based on the +/-16% bandwidth, which consistently exhibit the smallest standard errors.

For overall pass rates in math in 2014, the 95% confidence interval (CI) of this preferred estimate is [-10.355, 0.139], and in reading in 2014 it is [-6.871, 0.421]. Thus, while we cannot rule them out, any positive effects on pass rates are likely to be small. With respect to the gender subgroups, of note are the consistently large negative effects in math for males in both 2013 and 2014 and reading for females in 2013. Other subgroup effects are mixed, with some evidence of negative effects for black students in math in 2013 and reading in 2014. For Hispanic students, er find some evidence of negative effects in math in 2014 and reading in 2013. Many of the estimates are not statistically significant at traditional levels, which means we cannot rule out small positive effects for some of the subgroups. Nonetheless, given the many negative coefficients in the table, we can be quite confident in ruling out the hypothesis that the program had large positive effects, either overall or for any of the subgroups.

Moreover, we can rule out the possibility that any negative effects reflect prior year trends by extending the preferred analysis back in time to 2006, as shown in Figure 3. In the subgroup of schools that were open from 2006 through 2014, we find statistically significant negative effects in the overall composite score in 2014, in math in 2013 and 2014, and in reading in 2014. For the subgroup of schools open from 2006-2014, the 2014 95% CI is [-11.492, -2.717] in math and [-7.721, -1.057] in reading. We find no evidence of placebo effects in 2006 through 2009.

To supplement our analysis of how the program affected passing rates in the treated schools, we also explore how it affected student behavior (see bottom part of Table 2). We find that the TALAS program decreased average daily attendance by point estimates of 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, with point estimates 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 at least no better, student outcomes than comparable untreated schools.

Next, we turn to the student-level longitudinal analysis of students who had been in schools just below or just above the cut point in 2010. We limit the analysis to students who were in their third grade year in 2010. 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 show that the program had no observable overall effect on the passing rates of the treated students in either math or reading, where 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 11.3 to 21.0 percentage points in 2012, when most of them were in fifth grade. These are matched with a 0.127 to 0.289 SD increase in test scores for this group. 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. Following 2012, many of the students moved to middle schools that were not turnaround schools, and the gains faded as the students continued to progress through school (full results not shown). The passing rate point estimates for the Level II students in 2013 (when most would have been sixth graders) range from 2.2 (SE=4.2) to 4.6 (SE=2.9) percentage points. The magnitude and precision, though not the direction, of these estimates are sensitive to our choice of bandwidth. Hence the initial positive effects on level II students in the treated schools appear to be short term effects at best.

At the same time, we find consistently large reductions (point estimates of 0.356 to 0.641 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 pass 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. Further, these negative effects continue into at least one additional year (full results not shown). The point estimates for the top category of reading students range from -0.321 SDs (SE=0.207) to -0.740 (SE=0.314) in 2013 (when most would have been sixth graders). The 2014 estimates (when most would have been seventh graders) are null.

In sum, the turnaround program did not increase average achievement at either the school or the student level, and there is some evidence that it reduced schoolwide pass rates and the passing rates of some groups of students.[15] Based on the student level analysis, the only students

[15] We find no evidence of a difference in treatment effects by whether the schools were in RttT Districts (results not shown).

that may have gained from the program were those who were just below passing in 2010, though these gains did not persist and were not consistent across specifications.

6. Explaining the Patterns

With our detailed data on principals, teachers, and students, we can explore several possible explanations for the test score results: principal and teacher turnover, teacher time use, school climate, and the concentration of disadvantaged students in TALAS schools.

Consistent with the heavy use of the transformation option, we find that school principals left the treated schools at higher rates than other schools during 2012, the first full year after the program was implemented, though the effect is not statistically significant (see Figure 4 and Table 4).[16] The effectiveness of removing a principal depends on whether the new principals are more effective than the departing principals. Table 4 shows that the program led to a higher proportion of principals with less than four years of experience by 2014.

We find an uptick in teacher turnover in the year after the increase in principal turnover. Turnover may have increased 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.[17] We find no change in the proportion of inexperienced teachers. Figure 5 verifies the principal and teacher turnover results did not reflect prior-year trends. The figure shows no effect in placebo pre-treatment years back to 2009, but a large effect in 2012 for principal turnover and in 2013 for teacher turnover.[18]

---

[16] 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).

[17] Several schools mentioned placing low-performing teachers on action plans in their 2012 annual report, with the intention to remove them if they did not achieve growth. Other schools mentioned an increase in teacher resignations in 2013 for teachers not meeting principal expectations (Department of Public Instruction, 2013b, 2014).

[18] Estimates differ in Table 4 and Figure 5 because Figure 5 uses a linear spline for all years.

We next examine teacher time use (Table 5 and Figure 6). Several identified activities were required as part of the transformation and turnaround models, but others were not. The most consistent 2012 findings emerge for professional development, supervisory duties, required committee or staff meetings, and required paperwork, each of which increased as a result of TALAS. An increase in professional development was expected because it was intended as a key component of TALAS. The increase by 2014 in communicating with parents and the community was also consistent with the TALAS program of promoting community involvement. TALAS also promoted the use of ongoing assessments to track student progress. Teachers spent more time delivering assessments in treated schools by 2014, but they did not change the time they spent using the results of these assessments. Some caution may be necessary for the 2014 results given the high teacher turnover in treated schools in 2013.

Table 6 reports effects on teachers' perceptions of school climate. Positive numbers indicate increases in satisfaction in treated schools in standard deviation units. Though turnaround models emphasize school leadership, TALAS had no effect on teachers' perceptions of the quality of their schools' leadership. Nor did it have much effect on teachers' perceptions of the quality of professional development or community involvement. Some hints of dissatisfaction with facilities and resources emerged in 2012 (95% CI for +/-16% estimate [-0.823, 0.087]), along with concerns about time pressures in 2014 (95% CI [-0.813, 0.015]).

Finally, we find evidence that TALAS led to 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 chance that FRL students changed schools was fairly constant across the cut point. However, non-FRL students were much less likely to remain in the same school if they were in a school assigned to treatment in 2010, relative to those students not in treated schools (p =0.009). In other words, more affluent students from treated schools were more likely to attend a different school two years later than their less affluent counterparts. Table 7 confirms the increase in the proportion of FRL students at the school level across all years and across all methods. The 2014 point estimates range from a 3.4 to 6.0 percentage point increase in the proportion of FRL students in treated school.[19] There is no effect for the percentages of black or Hispanic students.

7. 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 specify 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 were 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 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.

One possible remaining concern is that other programs that were operating in North Carolina during this time could have affected our estimates, but only if their uptake was discontinuous at the TALAS cutoff point. For example, as noted NCDPI operated a district

[19] Non-movers (i.e., stayers), on average, were higher-achievers in the baseline 2010 year, scoring 0.192 SD higher in math and 0.162 SD higher in reading than movers. After controlling for school-level baseline scores (the running variable), this advantage remains large in control schools (e.g., 0.170 SD in math). However, the "stayer advantage" in baseline scores for the treated schools was much smaller (e.g., 0.087 SD in math). This implies that the leavers were more-advantaged in the turnaround schools, relative to leavers in the non-turnaround schools.

turnaround program during this period. In addition, NCDPI's Federal Programs division operated programs required by the Elementary and Secondary Education Act (Department of Public Instruction, 2015). Interviews with NCDPI staff indicated that the transformation division and Federal Programs division were distinct, with Federal Programs focusing on monitoring and TALAS on coaching. Nonetheless some of the projects under Federal Programs targeted schools that were also part of the TALAS program. In analysis shown in Appendix C, we examine whether there was an increase in the probability of assignment to these programs at the TALAS cut point, 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 school-level TALAS program in the RD design.

8. Conclusion

We find 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. Indeed, most of our estimated coefficients are consistent with the conclusion that the program reduced school wide pass rates and reduced the rates for some subgroups such as female students in math and male students in reading. Moreover, we show that the program affected the mix of students in the treated schools. The resulting greater concentration of low-income students in the treated schools could account for some of the disappointing findings at the school level. At the student level, the program may have led to higher scores and pass rates for the third grade students who were on the borderline of passing in math in 2010, but the improvements were short-lived. The program also may have reduced the test scores of the highest-achieving students in reading.

Hence, our results provide strong causal evidence against expanding the TALAS program at the margin. This conclusion contrasts with the implications of other recent research showing positive results for the same program (Henry et al., 2014, 2015). That research was based on a difference-in-difference (DID) analysis with the positive results largely driven by positive effects for the lowest performing schools.

Figure 2, described earlier, provides a visual depiction of the differing conclusions from the two methodologies. In particular, it shows null to negative differences across treated and control schools near the cut point (RD) but null to positive differences if the changes are averaged across the full range of data (DID). Our RD design prevents us from making strong causal conclusions about the effectiveness of the program for the very low-performing schools well below the eligibility cut point.

The difference between the RD estimates and the DID estimates could be caused by (at least) three factors that are not mutually exclusive. First, TALAS could have been more effective for the lowest-achieving schools. Second, changes to the test led to passing rates dropping in all schools from baseline to the post-intervention years. The lowest-achieving schools may have hit a floor, limiting their ability to drop further. Third, differences in outcomes at the lower end of the test score distribution could have been driven by the continuing effects of prior and concurrent interventions. As we document in Table 1, and discuss in detail in Appendix C, many of the treated elementary and middle schools also received prior state turnaround and other programmatic interventions. Because those other interventions were not based on the same eligibility requirements as TALAS, they would not interfere with our RD findings. They would, however, muddy the interpretation of DID models. That problem is exacerbated by the fact that the earlier studies included in the analysis not just the treated elementary and middle schools, but also the TALAS high schools, many of which were the target of major state interventions in prior years. Based on the reasonable assumption that turnaround programs and other state interventions may take several years to generate effects, we believe the DID strategy in the earlier research did not successfully isolate the effects of the TALAS program. In contrast, by measuring effects close to the cut point, our RD results are not driven by schools at the bottom of the distribution, which were most likely to have received multiple interventions. Robustness checks confirm our results; for instance, shrinking the bandwidth decreased sample size and increased standard errors, but did not change the direction of treatment effects.

The availability of North Carolina's biannual Teacher Working Conditions Survey allows us to open the black box to examine how teacher activities changed under a turnaround regimen. We conclude first that substantial changes 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, increased teacher turnover, and distracted teachers, potentially reducing the time available for instruction. 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. The mixture of principal replacement, teacher turnover, 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 may 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 was based clearly emphasized the need for continuity (Thomson, Brown, Townsend, Henry, & Fortner, 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.[20] Moreover, disadvantaged students need effective teachers and within-school structures of academic and social support to succeed. We find 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. Rural schools in particularly may require different staffing strategies than other school types. One potential lesson from the 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.

[20] Certain schools' annual reports mentioned 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 used backpack programs to provide food over the weekend for low-income children. However, because schools designed their own programs, these were 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.

Acknowledgements

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. Author 1 thanks Northwestern University's Institute for Policy Research and Author 2 thanks the Center for the Analysis of Longitudinal Data in Education Research funded by the U.S. Department of Education (#R305C120008) for financial support.

Works Cited

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), 444–455. https://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). New York, NY: 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. https://doi.org/10.1037/h0027982

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. https://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. https://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. https://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. (2012, June 13). ESEA Waiver Implications. Retrieved from http://www.ncpublicschools.org/docs/program-monitoring/esea/waiver.pdf

Department of Public Instruction. (2013a). *APR Report 2013 on Status of RttT TALAS Implementation by School* (RttT TALAS Report). Raleigh, NC. 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/

Dragoset, L., Thomas, J., Hermann, M., Deke, J., James-Burdumy, S., Graczewski, C., … Wei, T. (2016). *Race to the Top: Implementation and Relationship to Student Outcomes* (NCEE No. 2017–4001) (pp. 1–267). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/pubs/20174001/pdf/20174001.pdf

Dragoset, L., Thomas, J., Hermann, M., Deke, J., James-Burdumy, S., Graczewski, C., … Wei, T. (2017). *School Improvement Grants: Implementation and Effectiveness* (NCEE No. 2017–4013) (pp. 1–419). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/pubs/20174013/pdf/20174013.pdf

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. https://doi.org/10.1111/1468-0262.00183

Hansen, M. (2013). *Investigating the role of human resources in school turnaround: Evidence in two states* (Working Paper No. 89). National Center for Analysis of Longitudinal Date in

Education Research. Retrieved from http://www.caldercenter.org/sites/default/files/conferences/6th/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. https://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

Henry, G. T., Guthrie, J. E., & Townsend, L. W. (2015). *Outcomes and impacts of North Carolina's inititative to turn around the lowest-achieving schools* (pp. 1–56). Chapel Hill, NC: Consortium for Educational Research and Evaluation - North Carolina. Retrieved from http://cerenc.org/wp-content/uploads/2015/09/0-FINAL-Final-DST-Report-9-3-15.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

Hill, H. C. (2007). Learning in the teaching workforce. *Future of Children*, *17*(1), 111–127.

Imbens, G. W., & Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 933–959. https://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. https://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. https://doi.org/10.3102/0162373711398128

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

North Carolina Race to the Top Application. (2010). *Race to the Top Application*. Raleigh, NC: Office of the Governor. Retrieved from https://www2.ed.gov/programs/racetothetop/phase2-applications/north-carolina.pdf

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

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). New York, NY: Routledge.

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

Tang, Y., Cook, T. D., & Kisbu-Sakarya, Y. (2015). *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: An example from education*. Society for Research on Educational Effectiveness.

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. https://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. https://doi.org/10.1002/pam.21721

Appendix A: 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 the instructional practices construct, treated teachers were particularly dissatisfied with local assessment data being available in time to impact instructional practices in 2014. Within the time use construct, treated teachers were 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).

Appendix B: Details on the Estimation Strategy

We provide additional details on our estimation strategies in the following sections.

B.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. 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).

B.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 et al., 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) $$Turnaround_s = \alpha\, I(A_s \leq 0) + f(A_s) + \gamma X_s + \nu_s$$

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

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

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. Our models use the simplest model not rejected by this test; the vast majority have a linear spline on either side of the cutoff.

B.3 Parametric Analysis – Student-Level Analysis

For our analysis of the effects on student-level test scores, we use longitudinal data for individual students who were in third grade in a school +/-16 percentage points from the cut point in 2010, and limit the outcome variables to the year 2012. For our analysis of how the program affects the composition of students within a school, we use data for students in both 3rd and 6th grades 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 through 2014 in order to follow students as they move between available public schools.

The first stage predicts the probability of the student's 2010 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 can 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. 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.

Appendix C: Discontinuities in Simultaneous Programs

Additional programs could have affected the schools during the study period, including the original North Carolina school turnaround efforts, district-level RttT District turnaround, and Elementary and Secondary Education Act (ESEA) programs operated by NCDPI's Federal Programs division. The worry with these programs is that they may differentially occur on either side of the RD cut point.

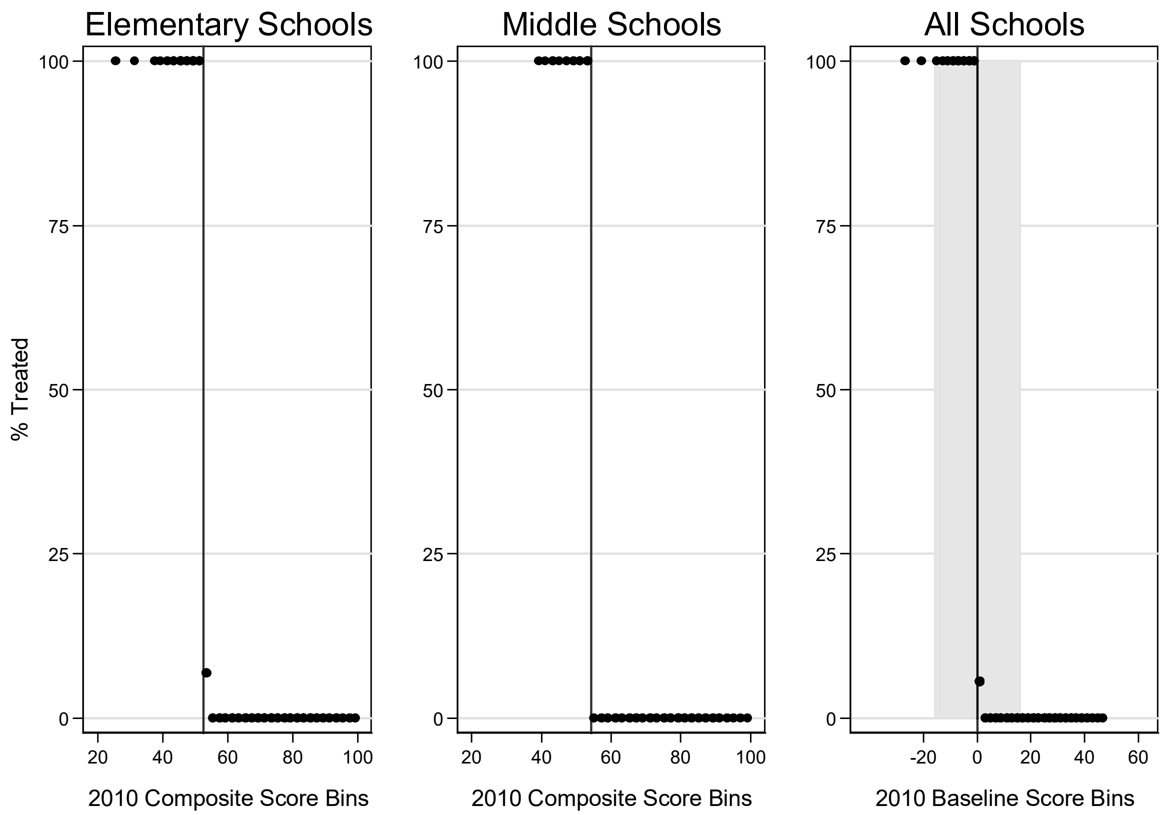
There is no jump in assignment to the original school turnaround program or RttT District Turnaround at the cut point (see Figure C.1). However, schools well below the cut point were more likely to be in these programs, which cautions against using difference-in-difference (DID) approaches.

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. Schools were assigned to their ESEA distinction using 2011 data, and schools remained in their category from the 2013 through 2015 school years. The assignment decision was announced at the end of the 2012 school year, and thus would not have affected our 2012 results (Department of Public Instruction, 2012). Moreover, to affect our 2013 and 2014 estimates there would have to be a difference in the ESEA program assignment at the 2010 TALAS cutoff. This is unlikely, because TALAS and ESEA schools do not have the same assignment mechanism. Assignment to an ESEA distinction was based on different years and either growth or absolute scores. Indeed, we find no statistically significant relationship between these programs at the cut point (see Figure C.1). 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/Focus.

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, though it cautions against using DID strategies.

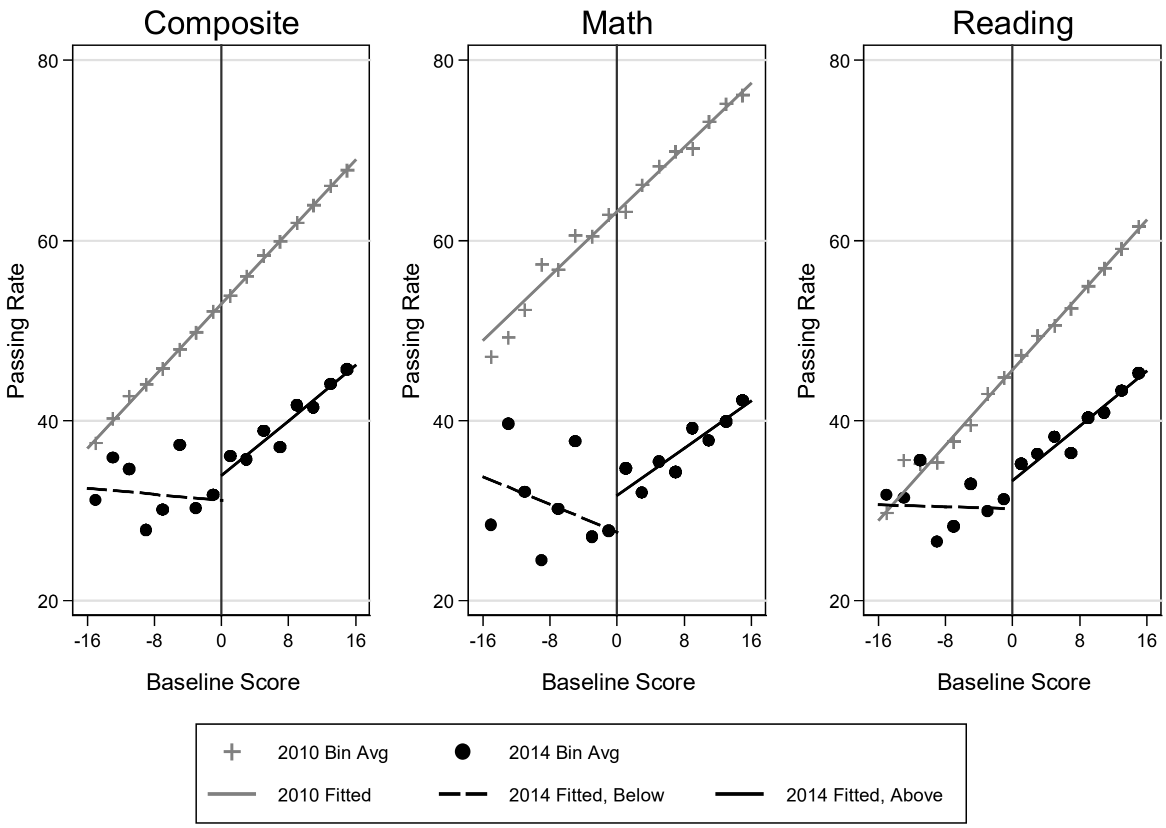
Figures

Figure 1: Treatment Uptake by School Type



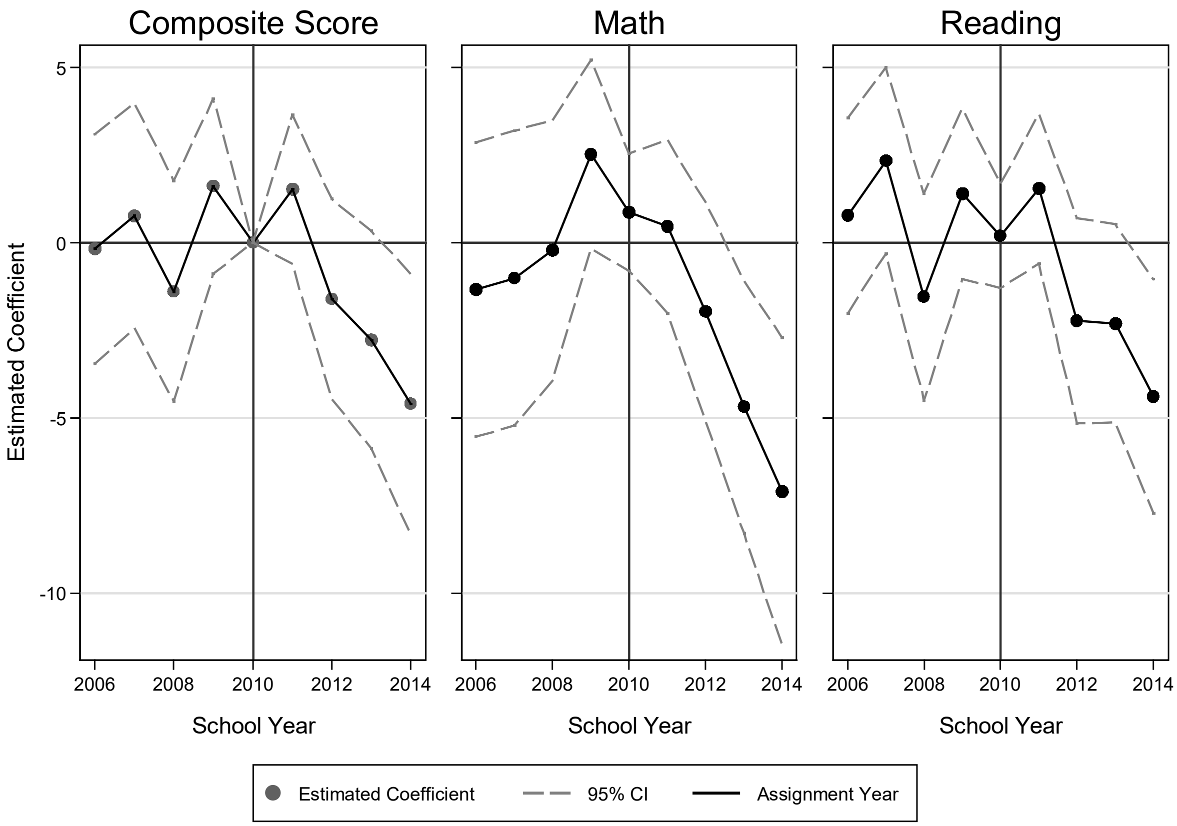
Note: Charts display the average uptake within 2.0 percentage point bins. Line indicates 2010 composite score cutoff. Grayed area indicates +/-16% from baseline cutoff.

Figure 2: 2014 Composite, Math, and Reading Scores



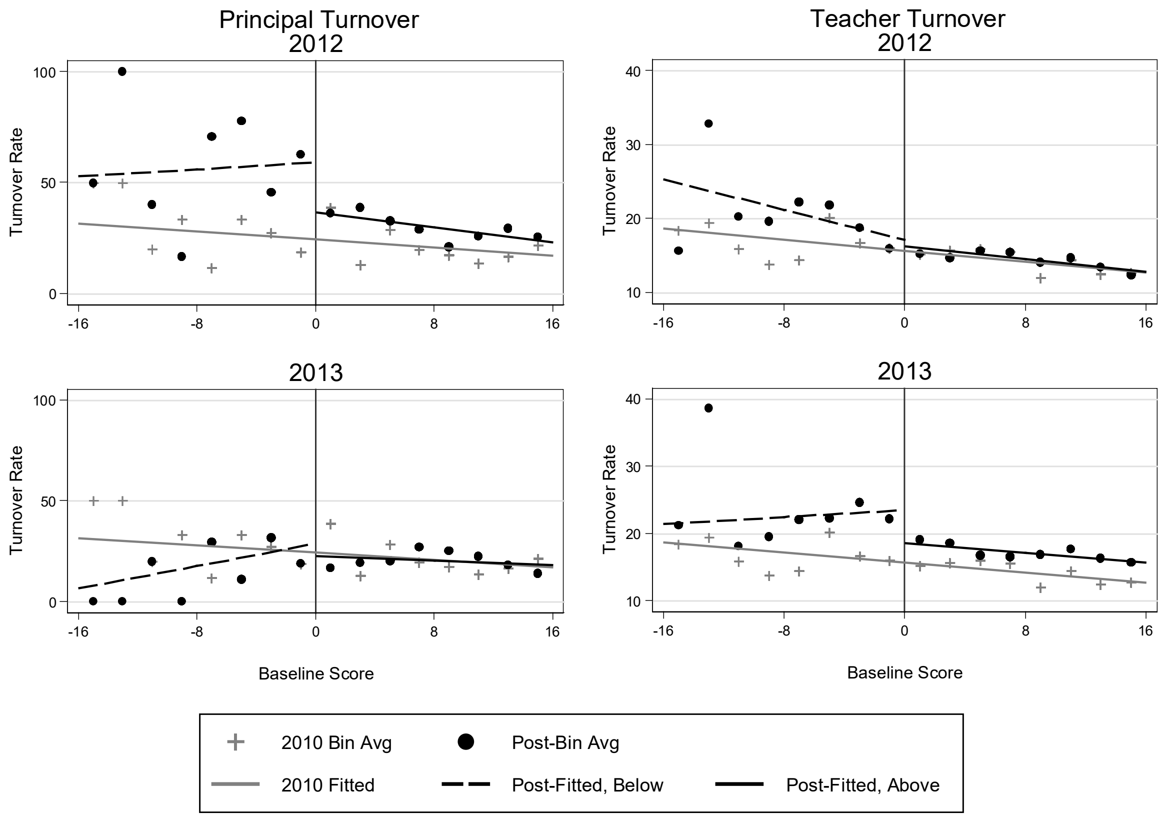
Note: Estimates of outcomes in 2010 and 2014 within +/-16% using our linear spline model with no additional controls (N=518 schools). Untreated post-period segment not constrained to be parallel with pre-period segments. All scores dropped from 2010 to 2014 due to a change in testing. Displayed bin width=2-percentage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

Figure 3: Test Results by Year



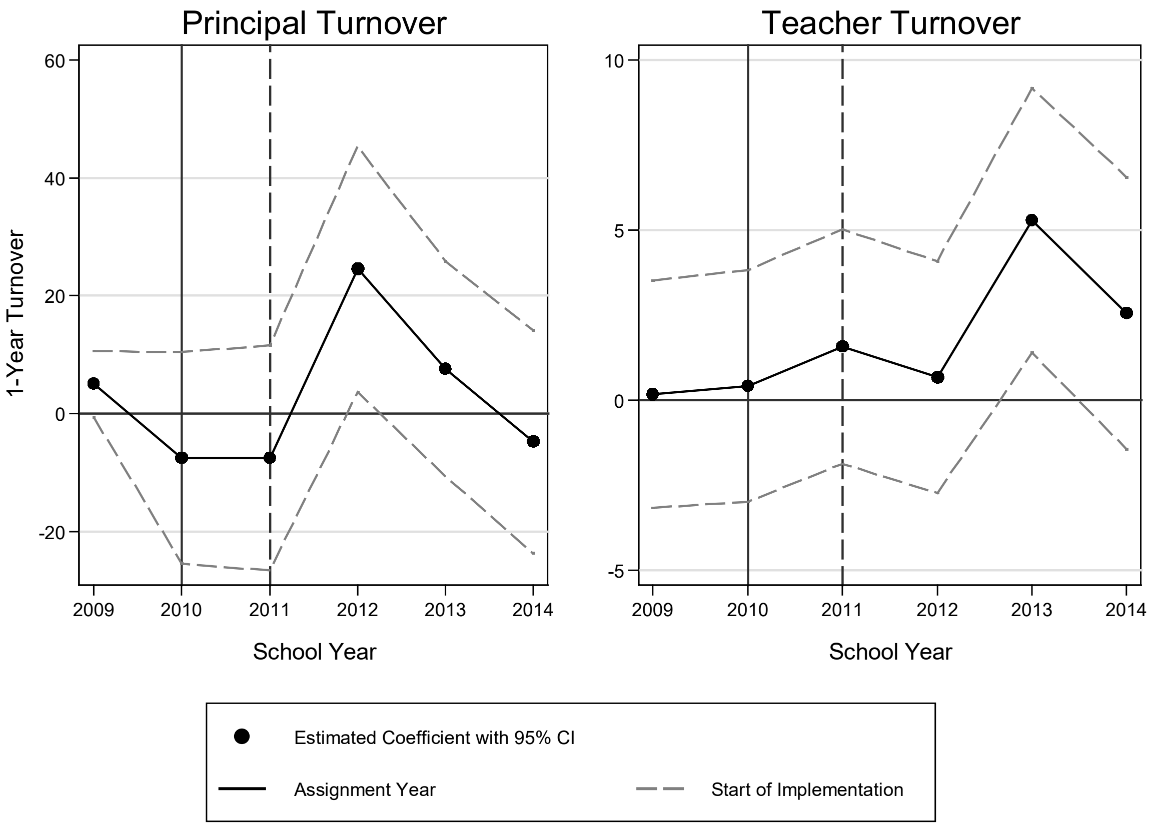
Note: Based on a separate +/-16% linear spline estimate with the same controls as Table 2; Year 2010 excludes baseline scores due to collinearity with the outcome. Only includes schools that appear in all years 2006-2014 (N=493 schools per year) to avoid compositional effects from schools that closed or opened over the period.

Figure 4: 2012 and 2013 Principal and Teacher Turnover



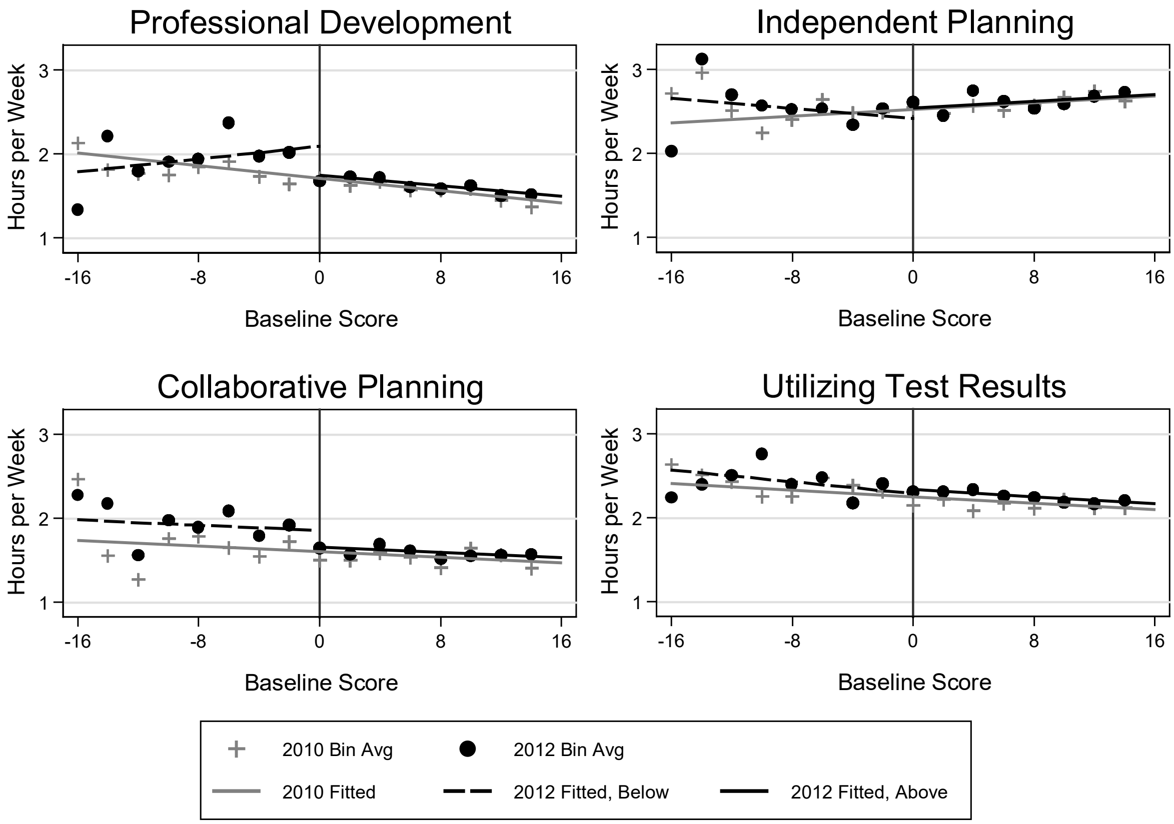
Note: Estimates of outcomes in 2010 and 2014 within +/-16% using our linear spline model with no additional controls (N=518 schools). Untreated post-period segment not constrained to be parallel with pre-period segments. Displayed bin width=2-percentage points.

Figure 5: Staff Turnover by Year



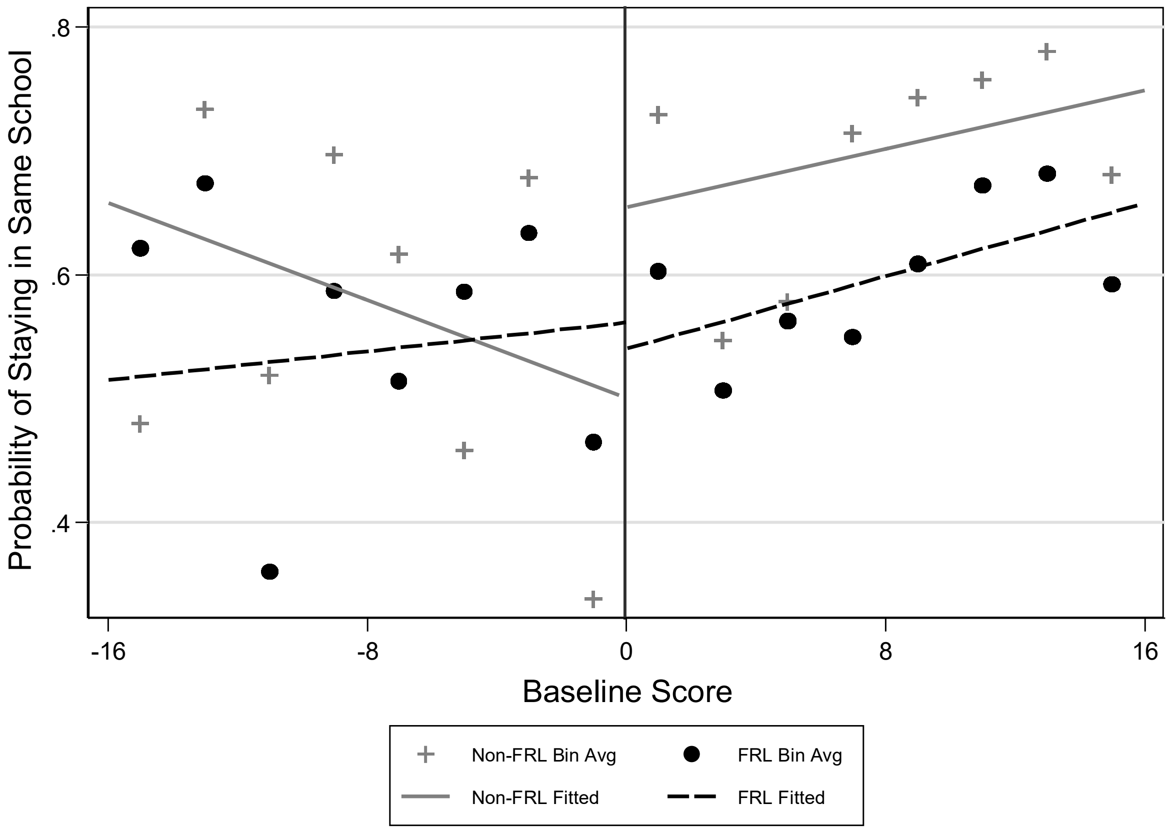
Note: Based on a separate +/-16% linear spline estimate with no additional controls for each year. Only includes schools that appear in all years 2009-2014 (N=512 schools per year) to avoid compositional effects from schools that closed or opened over the period.

Figure 6: 2012 Hours Spend on Activities per Week



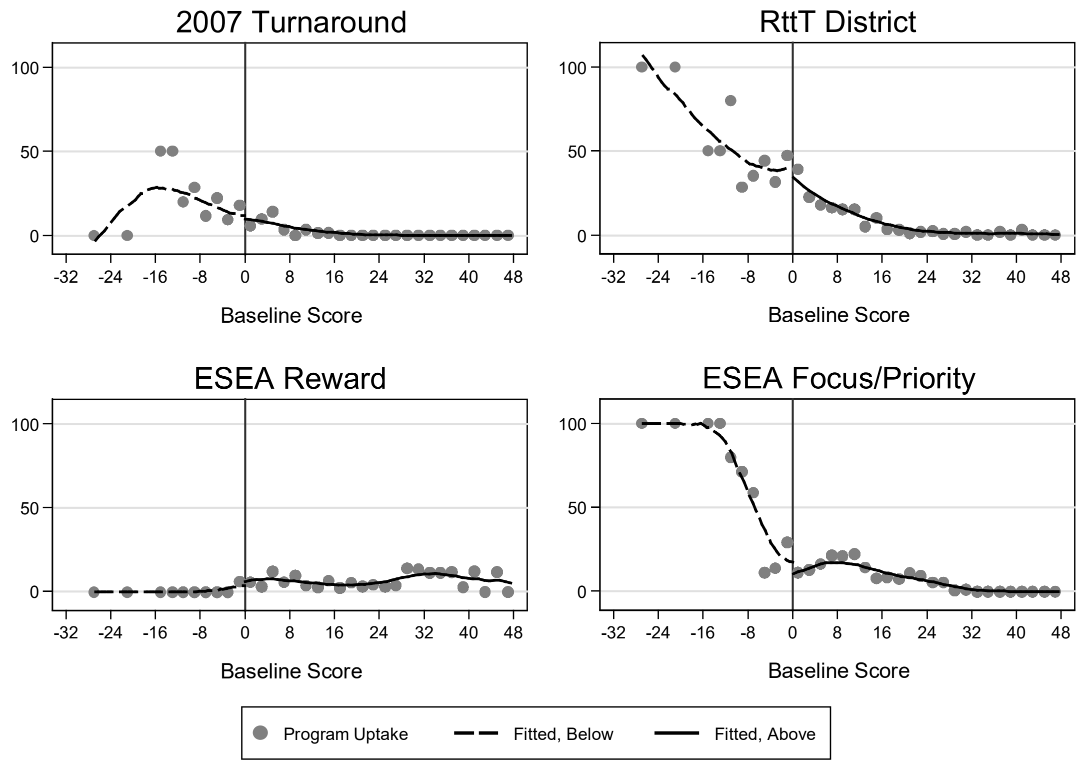
Note: Estimates of outcomes in 2010 and 2014 within +/-16% using our linear spline model with no additional controls (N=518 schools). Untreated post-period segment not constrained to be parallel with pre-period segments. Displayed bin width=2-percentage points.

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 (N=51,954 students). Displayed bin width=2-percentage points.

Figure C.1: Uptake of ESEA Reward/Priority/Focus Schools



Note: Nonparametric estimates based on 100% IK bandwidth. Displayed bin width=2-percentage points."2007 Turnaround" is the original turnaround program that the TALAS treatment was based on. "RttT District" is the district-level TALAS treatment based on 2011 test results. "ESEA Reward" is the 2012 assignment to the ESEA Reward designation. "ESEA Focus/Priority" is the 2012 assignment to either the ESEA Focus or ESEA Priority designation. All of these programs came with potentially different treatment than the school-level TALAS treatment based on 2011 test results.

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]		
	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	25.316	20.501	0.336		22.484	27.466	0.618
	(4.923)	(1.929)			(9.979)	(4.851)	
Principals w/ 0-3 Yrs. Exp.	43.038	42.597	0.942		45.006	45.662	0.953
	(5.606)	(2.363)			(11.095)	(5.527)	
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)	
Percent in Original NCDPI	16.456	4.100	0.000	***	9.547	10.098	0.945
Turnaround Program	(4.198)	(0.947)			(8.050)	(2.809)	
Percent in RttT Districts	41.772	15.490	0.000	***	37.849	30.666	0.510
	(5.584)	(1.729)			(10.897)	(4.716)	
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	0.928***	0.976***	0.960***	0.979***	0.976***	0.960***	0.949***	0.976***	0.960***
	(0.050)	(0.017)	(0.028)	(0.015)	(0.017)	(0.028)	(0.035)	(0.017)	(0.028)
F-Statistic	N/A	766,377	156,696	N/A	766,377	156,696	N/A	766,377	156,696
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)
Female 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)
Male 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)
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)
Female 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)
Male 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)		0.352	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 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	17.765	23.292[q]	20.433	9.993	9.654[q]	12.766	-5.312	-4.687	-3.464
	(11.013)	(16.514)	(13.682)	(10.803)	(13.452)	(11.260)	(11.055)	(9.917)	(12.061)
Principals with 0-3 Years of Exp.	-0.738	-2.406	-1.083	15.812	23.306*	24.394+	31.589*	27.707*	32.437*
	(11.961)	(11.433)	(14.168)	(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	0.276	0.537+	0.385***	0.311*	0.546*	0.486*[c]	0.101
	(0.199)	(0.280)	(0.114)	(0.139)	(0.260)	(0.206)	(0.128)
Individual planning	0.203	-0.129	0.045[c]	-0.238	0.296	-0.169	-0.144
	(0.388)	(0.269)	(0.368)	(0.188)	(0.372)	(0.174)	(0.211)
Collaborative planning	1.263***	0.556*	0.186	0.163	1.025***	0.023[q]	0.045
	(0.334)	(0.260)	(0.115)	(0.148)	(0.296)	(0.164)	(0.129)
Utilizing results of assessments	0.377	0.642*	-0.096	-0.163	-0.072	-0.096	0.052
	(0.449)	(0.280)	(0.115)	(0.154)	(0.241)	(0.115)	(0.145)
Administrative Burdens							
Supervisory duties	-0.098	0.332	0.421*[q]	0.270+	0.176	0.073	0.122
	(0.327)	(0.327)	(0.191)	(0.155)	(0.164)	(0.106)	(0.125)
Required committee/staff meetings	0.211	0.103	0.369**	0.288+	0.761***	0.343**	0.257+
	(0.238)	(0.275)	(0.125)	(0.156)	(0.231)	(0.117)	(0.151)
Completing required paperwork	0.239	0.511**	0.309*[q]	0.224+	0.351*	0.001	0.476*[q]
	(0.189)	(0.187)	(0.167)	(0.130)	(0.164)	(0.106)	(0.187)
Community & Students							
Communicating with parents/community	0.364**	0.312	-0.038[q]	-0.079	0.609***	0.100	0.333+[q]
	(0.137)	(0.205)	(0.109)	(0.091)	(0.172)	(0.085)	(0.180)
Addressing student discipline	0.091	0.340	0.099	0.304[q]	0.682	0.282	0.675[q]
	(0.252)	(0.311)	(0.164)	(0.337)	(0.443)	(0.188)	(0.413)
Focusing on Tests							
Prep for federal, state, and local tests	0.316	0.893***	0.036	0.121	0.439*	0.053	0.139
	(0.364)	(0.270)	(0.141)	(0.181)	(0.214)	(0.145)	(0.173)
Delivery of assessments	0.163	0.720**	-0.028	-0.011	0.397	0.193+	0.606*[q]
	(0.354)	(0.238)	(0.099)	(0.138)	(0.311)	(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

	2010	2012			2014		
	Nonparametric(1)		2SLS(2)		NP(1)	2SLS(2)	
Bandwidth	Varies	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%
Leadership	0.521	-0.447	-0.160	-0.198	-0.149	-0.088	-0.168
	(0.496)	(0.651)	(0.238)	(0.323)	(0.414)	(0.247)	(0.277)
Instructional Practices	1.296**	-0.044	0.087	-0.104	-0.236	-0.277	-0.334
	(0.459)	(0.372)	(0.207)	(0.273)	(0.381)	(0.227)	(0.261)
Professional Development	0.851*	-0.486	-0.164	-0.341	-0.073	-0.537+[q]	-0.398
	(0.416)	(0.416)	(0.253)	(0.327)	(0.340)	-0.298	(0.257)
Community Involvement	0.195	-0.488	-0.086	-0.172	-0.586+	-0.157	-0.217
	(0.423)	(0.489)	(0.207)	(0.264)	(0.341)	(0.193)	(0.248)
Student Conduct	0.509	-0.292	0.035	0.131	-0.251	-0.140	-0.088
	(0.440)	(0.414)	(0.221)	(0.278)	(0.500)	(0.241)	(0.281)
Facilities & Resources	0.309	-0.884*	-0.368+	-0.566*	-0.248	-0.265	-0.276
	(0.404)	(0.479)	(0.232)	(0.301)	(0.372)	(0.220)	(0.263)
Time	0.546	-0.505	-0.251	-0.517+	-0.221	-0.399+	-0.479+
	(0.404)	(0.479)	(0.232)	(0.301)	(0.420)	(0.211)	(0.261)
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

+ 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 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%	+/-10%	Varies	+/-16%	+/-10%	Varies	+/-16%	+/-10%
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	-0.276^{q}	-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 A.1: Survey Items and Factors

Construct	Question	2012, +/-16%	2012, +/-10%	2014, +/-16%	2014, +/-10%
School Leadership	Teachers are recognized as educational experts.	-0.073 (0.063)	-0.096 (0.084)	-0.057 (0.069)	-0.074 (0.080)
	Teachers are trusted to make sound professional decisions about instruction.	-0.072 (0.069)	-0.094 (0.092)	-0.049 (0.080)	-0.036 (0.093)
	Teachers are relied upon to make decisions about educational issues.	-0.069 (0.063)	-0.094 (0.082)	-0.036 (0.070)	-0.036 (0.084)
	Teachers are encouraged to participate in school leadership roles.	-0.023 (0.053)	-0.058 (0.070)	0.002 (0.053)	-0.025 (0.058)
	The faculty has an effective process for making group decisions to solve problems.	0.000 (0.073)	-0.019 (0.098)	-0.024 (0.075)	-0.040 (0.085)
	In this school we take steps to solve problems.	-0.011 (0.070)	-0.032 (0.094)	-0.037 (0.078)	-0.064 (0.090)
	Teachers are effective leaders in this school.	-0.036 (0.056)	-0.038 (0.076)	-0.016 (0.065)	-0.032 (0.073)
	Teachers have an appropriate level of influence on decision making in this school.	-0.069 (0.067)	-0.043 (0.091)	-0.045 (0.072)	-0.103 (0.079)
	The faculty and staff have a shared vision.	-0.022 (0.068)	-0.010 (0.094)	-0.029 (0.073)	-0.080 (0.081)
	There is an atmosphere of trust and mutual respect in this school.	-0.063 (0.088)	-0.056 (0.123)	-0.037 (0.094)	-0.069 (0.105)
	Teachers feel comfortable raising issues and concerns that are important to them.	-0.044 (0.091)	-0.031 (0.123)	0.042 (0.091)	0.022 (0.104)
	The school leadership consistently supports teachers.	-0.038 (0.085)	-0.022 (0.116)	0.014 (0.090)	-0.023 (0.100)
	Teachers are held to high professional standards for delivering instruction.	-0.004 (0.044)	-0.028 (0.057)	-0.053 (0.054)	-0.069 (0.063)
	Teacher performance is assessed objectively.	-0.038 (0.064)	-0.062 (0.084)	-0.002 (0.069)	-0.014 (0.077)
	Teachers receive feedback that can help them improve teaching.	-0.031 (0.068)	-0.093 (0.088)	-0.019 (0.076)	-0.080 (0.086)
	The procedures for teacher evaluation are consistent.	-0.055 (0.074)	-0.107 (0.093)	-0.072 (0.085)	-0.119 (0.090)
	The school improvement team provides effective leadership at this school.	-0.058 (0.068)	-0.077 (0.093)	-0.044 (0.067)	-0.089 (0.075)
	The faculty are recognized for accomplishments.	-0.028 (0.075)	-0.052 (0.099)	0.014 (0.069)	-0.048 (0.080)
	The school leadership makes a sustained effort to address teacher concerns about: Leadership issues	-0.036 (0.072)	-0.034 (0.098)	-0.051 (0.072)	-0.064 (0.083)
	The school leadership makes a sustained effort to address teacher concerns about: Facilities and resources	-0.031 (0.060)	-0.058 (0.079)	-0.053 (0.065)	-0.087 (0.073)
	The school leadership makes a sustained effort to address teacher concerns about: The use of time in my school	-0.047 (0.069)	-0.059 (0.095)	-0.017 (0.073)	-0.033 (0.084)
	The school leadership makes a sustained effort to address teacher concerns about: Professional development	-0.099 (0.065)	-0.111 (0.087)	-0.073 (0.064)	-0.110 (0.074)

	The school leadership makes a sustained effort to address teacher concerns about: Teacher leadership	-0.016 (0.062)	-0.044 (0.083)	-0.062 (0.066)	-0.086 (0.078)
	The school leadership makes a sustained effort to address teacher concerns about: Community support and involvement	-0.033 (0.059)	-0.049 (0.079)	-0.024 (0.066)	-0.049 (0.073)
	The school leadership makes a sustained effort to address teacher concerns about: Managing student conduct	-0.018 (0.077)	0.018 (0.106)	-0.009 (0.079)	-0.014 (0.090)
	The school leadership makes a sustained effort to address teacher concerns about: Instructional practices and support	-0.044 (0.061)	-0.071 (0.082)	-0.048 (0.064)	-0.087 (0.075)
	The school leadership makes a sustained effort to address teacher concerns about: New teacher support	-0.019 (0.071)	0.012 (0.093)	-0.044 (0.076)	-0.016 (0.093)
	Teachers are encouraged to try new things to improve instruction.	-0.007 (0.048)	-0.040 (0.063)	-0.007 (0.046)	-0.015 (0.054)
	Teachers have autonomy to make decisions about instructional delivery (i.e. pacing, materials and pedagogy).	-0.069 (0.065)	-0.112 (0.086)	0.014 (0.061)	0.004 (0.075)
	Overall, my school is a good place to work and learn.	-0.056 (0.067)	-0.043 (0.091)	-0.062 (0.081)	-0.070 (0.095)
Professional Development	Sufficient resources are available for professional development in my school.	-0.009 (0.056)	-0.054 (0.068)	-0.041 (0.063)	-0.095 (0.068)
	An appropriate amount of time is provided for professional development.	-0.008 (0.054)	-0.084 (0.068)	-0.011 (0.057)	-0.052 (0.063)
	Professional development offerings are data driven.	0.017 (0.058)	-0.045 (0.075)	-0.015 (0.049)	-0.029 (0.059)
	Professional learning opportunities are aligned with the school's improvement plan.	-0.031 (0.047)	-0.062 (0.061)	-0.014 (0.051)	-0.074 (0.056)
	Professional development is differentiated to meet the individual needs of teachers.	-0.060 (0.066)	-0.066 (0.088)	-0.053 (0.068)	-0.114 (0.076)
	Professional development deepens teachers' content knowledge.	-0.024 (0.055)	-0.052 (0.073)	-0.043 (0.054)	-0.091 (0.063)
	Teachers have sufficient training to fully utilize instructional technology.	-0.093 (0.064)	-0.095 (0.084)	-0.025 (0.059)	-0.064 (0.066)
	Teachers are encouraged to reflect on their own practice.	-0.014 (0.043)	-0.031 (0.055)	-0.000 (0.044)	-0.028 (0.048)
	In this school, follow up is provided from professional development.	-0.033 (0.063)	-0.058 (0.083)	-0.064 (0.069)	-0.123+ (0.074)
	Professional development provides ongoing opportunities for teachers to work with colleagues to refine teaching practices.	-0.036 (0.059)	-0.047 (0.078)	-0.043 (0.056)	-0.081 (0.065)
	Professional development is evaluated and results are communicated to teachers.	-0.040 (0.068)	-0.083 (0.090)	-0.031 (0.063)	-0.054 (0.078)
	Professional development enhances teachers' ability to implement instructional strategies that meet diverse student learning needs.	-0.033 (0.054)	-0.064 (0.072)	-0.044 (0.053)	-0.083 (0.063)
	Professional development enhances teachers' abilities to improve student learning.	-0.050 (0.051)	-0.074 (0.069)	-0.047 (0.050)	-0.100+ (0.059)
Community-School Relations	Parents/guardians are influential decision makers in this school.	-0.062 (0.060)	-0.073 (0.085)	-0.080 (0.077)	-0.153 (0.094)

	This school maintains clear, two-way communication with the community.	-0.036 (0.056)	-0.029 (0.077)	-0.020 (0.064)	-0.039 (0.079)
	This school does a good job of encouraging parent/guardian involvement.	-0.047 (0.059)	-0.068 (0.080)	-0.033 (0.067)	-0.046 (0.081)
	Teachers provide parents/guardians with useful information about student learning.	-0.020 (0.035)	-0.020 (0.044)	-0.030 (0.043)	-0.031 (0.052)
	Parents/guardians know what is going on in this school.	-0.013 (0.056)	-0.017 (0.077)	-0.026 (0.061)	-0.026 (0.072)
	Parents/guardians support teachers, contributing to their success with students.	0.039 (0.054)	0.006 (0.073)	-0.007 (0.064)	-0.045 (0.080)
	Community members support teachers, contributing to their success with students.	0.039 (0.056)	-0.010 (0.078)	-0.101 (0.072)	-0.167+ (0.088)
	The community we serve is supportive of this school.	-0.024 (0.064)	-0.088 (0.085)	-0.063 (0.069)	-0.086 (0.091)
	Students at this school understand expectations for their conduct.	-0.037 (0.067)	-0.037 (0.087)	0.000 (0.076)	-0.001 (0.085)
Facilities & Resources	Teachers have sufficient access to appropriate instructional materials.	-0.101 (0.070)	-0.156+ (0.081)	-0.054 (0.073)	-0.078 (0.086)
	Teachers have sufficient access to instructional technology, including computers, printers, software and internet access.	-0.110 (0.088)	-0.187+ (0.109)	-0.024 (0.077)	0.004 (0.092)
	Teachers have access to reliable communication technology, including phones, faxes and email.	-0.089 (0.062)	-0.145+ (0.078)	-0.072 (0.063)	-0.080 (0.073)
	Teachers have sufficient access to office equipment and supplies such as copy machines, paper, pens, etc.	-0.056 (0.078)	-0.122 (0.096)	-0.069 (0.084)	-0.130 (0.095)
	Teachers have sufficient access to a broad range of professional support personnel.	-0.032 (0.054)	-0.035 (0.070)	-0.007 (0.056)	-0.027 (0.063)
	The school environment is clean and well maintained.	-0.091 (0.064)	-0.107 (0.091)	-0.058 (0.076)	0.014 (0.100)
	Teachers have adequate space to work productively.	-0.075 (0.054)	-0.123+ (0.069)	-0.074 (0.051)	-0.061 (0.065)
	The physical environment of classrooms in this school supports teaching and learning.	-0.092+ (0.053)	-0.106 (0.072)	-0.056 (0.054)	-0.051 (0.068)
	The reliability and speed of Internet connections in this school are sufficient to support instructional practices.	-0.076 (0.074)	-0.127 (0.095)	-0.114 (0.075)	-0.145 (0.089)
Student Conduct	Students at this school follow rules of conduct.	-0.051 (0.087)	-0.065 (0.117)	-0.060 (0.102)	-0.060 (0.125)
	Policies and procedures about student conduct are clearly understood by the faculty.	0.004 (0.061)	0.031 (0.083)	-0.043 (0.070)	-0.019 (0.082)
	School administrators consistently enforce rules for student conduct.	0.027 (0.101)	0.086 (0.131)	-0.035 (0.109)	-0.007 (0.127)
	School administrators support teachers' efforts to maintain discipline in the classroom.	0.037 (0.096)	0.059 (0.125)	-0.010 (0.095)	0.020 (0.109)
	Teachers consistently enforce rules for student conduct.	-0.007 (0.046)	-0.006 (0.061)	-0.066 (0.051)	-0.065 (0.064)
	The faculty work in a school environment that is safe.	-0.031 (0.059)	-0.009 (0.079)	-0.078 (0.070)	-0.058 (0.082)
Instructional Practices	The school leadership facilitates using data to improve student learning.	-0.031 (0.051)	-0.068 (0.066)	-0.051 (0.055)	-0.060 (0.066)

	State assessment data are available in time to impact instructional practices.	0.038 (0.046)	-0.025 (0.058)	-0.091 (0.056)	-0.106 (0.071)
	Local assessment data are available in time to impact instructional practices.	0.018 (0.046)	-0.019 (0.060)	-0.084+ (0.050)	-0.104+ (0.058)
	Teachers use assessment data to inform their instruction.	-0.001 (0.034)	-0.025 (0.048)	-0.055 (0.038)	-0.065 (0.044)
	Teachers work in professional learning communities to develop and align instructional practices.	-0.020 (0.048)	-0.037 (0.064)	-0.049 (0.050)	-0.035 (0.057)
	Provided supports (i.e. instructional coaching, professional learning communities, etc.) translate to improvements in instructional practices by teachers.	-0.012 (0.049)	-0.021 (0.063)	-0.026 (0.049)	-0.042 (0.056)
Time	Class sizes are reasonable such that teachers have the time available to meet the needs of all students.	-0.072 (0.086)	-0.119 (0.112)	-0.062 (0.090)	-0.054 (0.105)
	Teachers have time available to collaborate with colleagues.	-0.090 (0.074)	-0.162+ (0.095)	-0.105 (0.066)	-0.108 (0.080)
	Teachers are allowed to focus on educating students with minimal interruptions.	-0.044 (0.073)	-0.095 (0.095)	-0.147+ (0.083)	-0.186+ (0.099)
	The non-instructional time provided for teachers in my school is sufficient.	-0.103 (0.079)	-0.150 (0.107)	-0.129 (0.086)	-0.148 (0.097)
	Efforts are made to minimize the amount of routine paperwork teachers are required to do.	-0.068 (0.073)	-0.178+ (0.093)	-0.101 (0.076)	-0.173* (0.084)
	Teachers have sufficient instructional time to meet the needs of all students.	-0.033 (0.054)	-0.112+ (0.067)	-0.144* (0.061)	-0.174* (0.074)
	Teachers are protected from duties that interfere with their essential role of educating students.	-0.139* (0.068)	-0.201* (0.084)	-0.115+ (0.069)	-0.169* (0.077)
	Teachers are assigned classes that maximize their likelihood of success with students.	-0.002 (0.070)	-0.026 (0.094)	-0.024 (0.065)	-0.031 (0.077)