ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-107

The Effects of Race to the Top School Turnaround in North Carolina

Gary T. Henry
Vanderbilt University

J. Edward Guthrie
Vanderbilt University

Federal education policies gave political and financial support for state education agencies to turnaround low-performing schools on an unprecedented scale. North Carolina's ambitious program turned around over half of all schools nationwide that underwent turnaround funded by Race to the Top. Exploiting the assignment to turnaround based on schools' 2009-10 proficiency rates, we implement regression discontinuity designs to estimate the effects of state turnaround services on student achievement in North Carolina's lowest-performing schools annually from the 2011-12 through 2014-15. Overall, we find modest positive effects of turnaround when including treated schools at all grade levels, but these effects are sensitive to bandwidth. For secondary schools, we find consistently positive effects that vary from modest to large. For elementary and middle schools, we find consistent, modest negative effects of turnaround on student achievement.

VERSION: July 2019

Suggested citation: Henry, Gary T., and J. Edward Guthrie. (2019). The Effects of Race to the Top School Turnaround in North Carolina. (EdWorkingPaper: 19-107). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-107

The Effects of Race to the Top School Turnaround in North Carolina

Gary Henry

gary.henry@vanderbilt.edu

J. Edward Guthrie

j.edward.guthrie@vanderbilt.edu

Peabody College

Vanderbilt University

Corresponding author:

J. Edward Guthrie
j.edward.guthrie@vanderbilt.edu
(615) 584-1778

This manuscript was partially funded by the U.S. Department of Education through the Race to the Top grant to the North Carolina Department of Public Instruction.

ABSTRACT

Federal education policies gave political and financial support for state education agencies to turnaround low-performing schools on an unprecedented scale. North Carolina's ambitious program turned around over half of all schools nationwide that underwent turnaround funded by Race to the Top. Exploiting the assignment to turnaround based on schools' 2009-10 proficiency rates, we implement regression discontinuity designs to estimate the effects of state turnaround services on student achievement in North Carolina's lowest-performing schools annually from the 2011-12 through 2014-15. Overall, we find modest positive effects of turnaround when including treated schools at all grade levels, but these effects are sensitive to bandwidth. For secondary schools, we find consistently positive effects that vary from modest to large. For elementary and middle schools, we find consistent, modest negative effects of turnaround on student achievement.

INTRODUCTION

Whether attributed to increasing income inequality, resegregation, bureaucratic mismanagement, or failed social policies, failing schools have become a target of public concern as indicated by the media focus on “dropout factories” (see for example http://hosted.ap.org/specials/interactives/wdc/dropout/index.html) and educational policy initiatives aimed at school turnaround. Emphasis on “failing schools” as a primary unit of educational reform grew out of the school accountability movement of the 1990s (Murphy & Meyers, 2008). The school-level measures of academic achievement and attainment, which became ubiquitous as a result of *No Child Left Behind* brought chronic and concentrated under-performance in some schools, especially those serving high percentages of economically disadvantaged and minority students, to the attention of policymakers and the public (Balfanz & Legters, 2004). In the face of political resistance and legal restriction to undertake school integration or desegregation efforts (Rothstein, 2013), federal education reform has focused on what Rothstein (2013) terms “compensatory education,” targeting conditions and academic performance in low-achieving schools in a set of efforts collectively known as school turnaround (Murphy & Meyers, 2008).

Over the past several years, federal initiatives focused on school turnaround drew funds from the American Recovery and Reinvestment Act of 2009 for both School Improvement Grants (SIG) and the Race to the Top state competition. In addition to district applications for SIG, Race to the Top offered states the opportunity to compete for grants by proposing K-12 educational reforms to address areas identified as priorities by the U.S. Department of Education (ED). One of Race to the Top’s four main reform priorities was to turn around performance in low-achieving schools and districts. In the call for grant applications, ED specified that applicants must submit a plan to significantly improve performance in schools in the lowest five percent of student proficiency. ED further stipulated that Race to the Top and SIG reform efforts in these low-achieving schools follow one of four federally-approved turnaround models—transformation model, turnaround model, closure model, or restart model. Although the latter two models divest districts of traditional control either through dissolution of the school and reassignment of its students (closure) or transfer of school oversight to a charter management or educational management organization (restart), the former two prescribe specific staff replacement and structural reform features within the traditional district-school governance and management structure.

In 2010, ED awarded North Carolina $400 million to implement the state's proposal submitted for the federal Race to the Top program. To fulfill the school turnaround commitment of North Carolina's successful application for funding, the District & School Transformation division of the North Carolina Department of Public Instruction used a share of the state's Race to the Top funds to support Turning Around Low-Achieving Schools (TALAS). Beginning with the 2010-11 school year, TALAS reformed 106 schools under the transformation and turnaround models and closed 12 more (ED, 2014). The present study examines the effects of the school turnaround efforts implemented by TALAS, which required among other reforms: leadership change, staff replacement, increased efforts to recruit and retain effective educators, and coaching and other job-embedded professional development. Assignment of the lowest performing five percent of North Carolina schools to turnaround followed a sharp cutoff based on 2009-10 school-level proficiency rates (test-level pass rate for end-of-grade and end-of-course exams). This strict assignment to treatment creates an opportunity to estimate a local average treatment effect (LATE) using a regression discontinuity design (RDD), which exploits the exogenous threshold between treatment and control to support strong causal inferences from differences in outcomes between treated and untreated observations near the assignment threshold (Cook, 2008; Lee & Lemieux, 2009; Van der Klaauw, 2008).

Specifically, this study uses an RDD to estimate the impact of North Carolina's school turnaround efforts on student achievement as measured by state achievement exams from the 2011-12 school year through 2014-15.

BACKGROUND

ED first prioritized whole school reforms with the Comprehensive School Reform (CSR) program in 1998, noting a lack of research or understanding regarding the processes most successful in raising achievement in chronically low-performing schools (ED, 2001). The program awarded grants to low-performing schools with an expectation that these schools would implement comprehensive reform efforts consisting of "scientifically-proven" methods to address all aspects of school operation. However, even the most promising reform features were no more likely to be implemented in CSR reward schools than in other schools, and the reforms implemented generally lacked a scientific research basis (ED, 2008). Not surprisingly given this lack of evidence-based choices for CSR models and subsequent low implementation fidelity, CSR awards generally failed to improve student outcomes in award schools (Bifulco, Duncombe, & Yinger, 2005; Gross, Booker, & Goldhaber, 2009; ED, 2008) .

School accountability sanctions under No Child Left Behind (NCLB) beginning in 2003 also included prescriptive school reforms. These sanctions include forcing schools to submit a School Improvement Plan (SIP), allowing students to transfer to higher-performing schools, replacing leadership and staff, and reorganizing and restructuring the school. Like CSR, the reform prescriptions of NCLB were not enacted as intended and failed to produce their desired results. Specifically, SIPs have tended to be unworkable reform models with little faculty buy-in (Mintrop & MacLellan, 2002; Mintrop, MacLellan, & Quintero, 2001), few students' families have exercised choice options (Brownstein, 2003), and states and districts typically lacked evidence about the effectiveness of alternative reform strategies to guide their choice of strategies or the human capital to effectively reconstitute entire schools (Herman et al., 2008).

Turning around failing schools has been an increased priority of the ED's federal reform efforts under Secretary of Education Arne Duncan. Within six months of his speech in June 2009 on the importance of improving student outcomes in "chronically low-achieving schools" (Duncan, 2009), ED allocated funding from the American Recovery and Reinvestment Act of 2009 to state education agencies for school turnaround efforts through competitive grants under State Fiscal Stabilization Funds, Race to the Top, and increased support for a pre-existing but previously underfunded School Improvement Grant (SIG) program authorized under Title 1 of the Elementary and Secondary Education Act (ESEA) of 1965. In addition, in specifying the selection criteria for Race to the Top, ED included school turnaround among the reforms state education agencies must agree to enact in order to be awarded funding. This section of the application asked states to propose a plan to identify their lowest-achieving schools, demonstrate the authority and capacity to intervene, and commit to reforming each intervention school with one of the four department-approved turnaround models—the turnaround model, restart model, closure model, or transformation model. ED first expressed its belief in these models when it stipulated that schools select and follow one of the four models in order to receive SIG awards, and the department has consistently endorsed these turnaround models in subsequent reform initiatives. Unlike the reform prescriptions under CSR or NCLB sanctions, school turnaround under Race to the Top and SIG require commitments to substantial and immediate change at the school level under one of the four available turnaround models.

The closure model is the most extreme and disruptive of the four reform models, as the school is permanently closed, its students re-enrolled elsewhere in the district, and the leadership and faculty laid off with no guarantee of being rehired by another school. The restart model involves transferring management to an approved charter management organization (CMO) or education management organization (EMO). Restarts also require that schools allow any student enrolled in the school prior to restart to re-enroll in the new, independently-managed school. Under the turnaround and transformation models, schools remain open and under district management but undergo significant staff replacement and comprehensive reform. Both the turnaround and transformation models involve replacing the school principal, but differ in the amount of teacher replacement and number of other reforms required. Turnaround requires that in addition to replacing the principal, the LEA rehire no more than half of the school's staff from the previous year. Beyond the staff replacement, however, the principal has greater autonomy to enact reform measures. The transformation model does not require staff replacement beyond the principal, but requires the school to accept management oversight of the implementation of specific reforms including more rigorous and evidence-based teacher evaluation, job-embedded professional development, curricular overhaul, and increased learning time through extension of either the school day or school year. The least disruptive of the four models from both staffing and governance perspectives, transformation, has been the dominant choice for reform models, implemented in almost 70% of Race to the Top turnaround schools and over 70% of SIG schools nationwide (Dee, 2012; ED, 2011, 2014).

School turnaround efforts under Race to the Top grants are further distinct from SIG awards in that they call for state education agencies to oversee reform implementation rather than schools or districts. The 117 turnaround schools in North Carolina represented almost one-third of the 380 schools falling under state-led Race to the Top turnaround programs nationwide (ED, 2014), and thus present an important early opportunity to evaluate the potential for state education agencies to manage successful school turnaround initiatives.

Reviewers judged North Carolina's plan for school turnaround to have been a major strength in its application for Race to the Top funding, scoring 40 out of 50 possible points in this section of its grant application. The federal grant reviewers noted favorably the state's legal authority to intervene in districts and schools as well as a capacity for turnaround efforts demonstrated by its "Turnaround Schools" program that operated from 2006 to 2010. The District and School Transformation (DST) division of NCDPI administered the previous school transformation initiative, which was charged with improving proficiency and graduation outcomes of every school in the state below 60% in either measure (Thompson, Brown, Townsend, Henry, & Fortner, 2011; Thompson, Henry, & Preston, 2016). The state's new turnaround initiative under Race to the Top, known as TALAS (Turning Around Low-Achieving Schools), would operate under the DST and enact the federally prescribed reform models in the lowest five percent of schools in the state as measured by 2009-10 proficiency rates on required state tests[1].

The TALAS delivery model is one of direct service from the state's DST division. The DST deployed roughly 150 coaches to its treatment schools, including district transformation coaches who

[1] In North Carolina, students take end-of-grade (EOG) tests in reading and mathematics in grades 3-8, and end-of-course (EOC) tests in a number of courses usually taken during ninth or tenth grade. For each EOC and unique subject-grade EOG exam, cut scores demarcate the level of command a student demonstrates over the tested knowledge and skills as "Limited" (Level 1), "Partial" (Level 2), "Sufficient" (Level 3), "Solid" (Level 4), or "Superior" (Level 5). See Table 2 for a summary of exams included in the calculation of passing rates by year. Passing rates are calculated as the percentage of all EOC and EOG tests taken in a school each year which meet a standard of at least "Sufficient" (Level 3), the level demarking proficiency for school accountability under NCLB (North Carolina Department of Public Instruction, 2010). However, unlike NCLB reporting in which proficiency is reported as the percentage of students achieving proficiency by subject and grade, for schools' overall PR it is the total number of tests taken and not the number of students tested which serves as the denominator because individual students often take tests in multiple subjects each year, creating multiple student-level observations within schools. For 2,489 North Carolina schools in the 2009-2010 school year, the year used to determine assignment to TALAS, the average PR was 73.8%. By school level, the average passing rates were 73.8% for elementary schools, 70.9% for middle schools, and 76.9% for secondary schools.

worked with district personnel to devise and implement strategies to better support districts' lowest-performing schools; school transformation coaches who worked primarily with school-level administrators to improve leadership activities and align them with the school turnaround strategies; and instructional coaches whose work focused on curriculum and instruction, through one-on-one coaching with teachers, advising professional learning communities (PLCs) or other multi-teacher work groups, and targeted professional development. The majority of DST's coaches served in the instructional coaching role, meeting with teachers at each treatment school multiple times per week. School and district transformation coaches served more schools and met with worked with their assigned schools and districts less frequently than instructional coaches.

Existing and emerging research has begun to suggest that turnaround strategies can be effective in raising student achievement in chronically low performing schools. Dee (2012), who estimated credible causal effects employing a RDD, found that SIG-sponsored reforms in California produced positive effects, but additional analysis (difference-in-differences) suggested that these effects were concentrated in schools implementing the more disruptive turnaround model and not the transformation model. In a study of turnaround in three schools in a large urban district, Strunk, Marsh, Hashim, and Bush-Mecenas (2016) find initially positive effects in students' English/Language Arts scores (but not in math), but that positive effects were not sustained as the schools' emphasis on innovation and growth waned. Papay (2015) finds similar positive effects using RDD, comparative time series, and interrupted time series, in which schools labeled as the lowest-performing in Massachusetts improved student achievement initially and improved the trajectory of student achievement over four years. However, the Papay study is silent on the actions taken by these chronically low-performing schools to achieve these improvements. In a study of a district turnaround, Schueler, Goodman, and Deming (2016) detail the "portfolio approach" to turnaround used in Lowell, MA, and find positive gains in the initial year of operation after takeover. In a study comparing the effects of state takeover in which the management of most schools was placed in the hands of charter management operators (Achievement School District) to those of turnaround managed by local school districts (iZones), Zimmer, Henry, and Kho (2016) find that the iZone schools produced consistent moderate to large gains over their first three years of operation, while the Achievement School District schools have not produced consistent effects. Overall, these studies suggest that effects may be evident more quickly than some education scholars had hypothesized (Berends, Bodilly, & Kirby, 2002), but that the actions taken to produce gains in achievement are not well understood, and the sustainability of these gains is highly variable.

DATA

The following analyses draw from a comprehensive, longitudinal statewide dataset maintained by the University of North Carolina-Chapel Hill's Educational Policy Initiative at Carolina (EPIC). For student achievement, the sample includes student-level test scores on math and English end-of-course (EOC) for high school students and end-of-grade (EOG) achievement exams in reading and math for grades four through eight[2]. The analyses thus contain scores from four exams for typical elementary schools, six exams for middle schools, and two for high schools.

Covariate adjustments, including prior test scores, are made at the student level rather than at the school level to more precisely identify treatment effects on individual student performance. Additional covariates include binary indicators for student race, gender, ELL status, economic disadvantage, academic or intellectual giftedness, being either overage or underage, and mid-year

[2] Exam scores from third grade students are not included as outcome variables because of the absence of a prior score to serve as a control for students' natural endowments and prior inputs.

movers. The model also includes indicators for students in advanced and remedial tracks for the subject tested, prior achievement scores, and days absent during the current school year.

The primary independent variable is the assignment variable, the school-level proficiency rate for all state EOG and EOC exams for 2009-10. The denominator for this measure is the number of exams taken, rather than the number of students, as students take multiple tests across subjects and their proficiency is measured separately for each exam. Proficiency rates are standardized and centered at the threshold value of the forcing variable by level (elementary, middle, or secondary).[3]

TALAS began implementation midway through the 2010-11 school year, but was not fully implemented in all turnaround schools until 2011-12. We estimate the effect of the TALAS intervention in terms of student achievement gains in the 2011-12, 2012-13, 2013-14, and 2014-15 school years.

SAMPLE

The DST applied several exclusion criteria to determine eligibility before identifying schools for TALAS treatment. Schools ineligible for participation in TALAS included charter schools, state-run schools (without a local education agency), schools serving special populations of students such as students with disabilities or under correction or rehabilitation, and schools with non-traditional themes. Of the 2,756 schools with performance data for 2009-10, the exclusion criteria identify 307schools ineligible for turnaround assignment and 2,449 eligible. We apply these categorical

[3] Because values of the forcing variable vary systematically by grade level, the DST identified the bottom five percent of elementary, middle, and secondary schools separately. Because of the systematically higher proficiency rates of high schools, simply taking the lowest 5% of all schools would have resulted in an over-representation of elementary schools and an under-representation of high schools in the treatment group. Treatment within each level followed strict assignment with 2009-2010 proficiency rate acting as the forcing variable. Due to the separate identification of the lowest five-percent of schools by level, the numeric cut-point for assignment varies between elementary, middle, and secondary schools. For elementary schools, the cut-point is 52.4%; for middle schools 53.3%; and for high schools 58.1%. These thresholds forced an assignment to treatment of 65 elementary schools (4.80% of eligible elementary schools), 24 middle schools (5.35%), and 19 high schools (4.11%).

exclusion criteria to restrict the sample to only those schools potentially eligible for turnaround intervention. Based on the 2009-10 proficiency rates, DST assigned 110 schools to turnaround.

Additionally, the TALAS treatment group includes seven high schools above the proficiency rate threshold that were selected for turnaround based a second forcing variable threshold, the graduation rate and cutoff below 60%. One of these schools closed before service implementation, while the other six followed the transformation model. Of the four schools initially identified as below both the graduation and proficiency assignment thresholds, three were separate academies within the same building and closed before service implementation. The inclusion of the six high schools above the proficiency threshold but below the graduation threshold is significant for the design of this study because it compromises the otherwise strict assignment based on the primary forcing variable. We use two strategies to address these crossover schools— "fuzzy" RDD and frontier RDD. These methods are discussed further in the Methods section.

Another potential confounder exists in the concurrent SIG awarded to 24 North Carolina schools based on similar performance criteria to the TALAS assignment (US Department of Education, 2011). Though SIG recipients follow identical reform models to those implemented by the DST, their turnaround efforts fall under the purview of their districts and not the DST office of NCDPI. For this reason, we do not consider them as treatment schools in this evaluation. SIG schools present minimal threat to bias the effect estimates of this study as a majority of the grants were awarded to alternative schools ineligible for TALAS due to the exclusion criteria detailed above. Of the 24 SIG schools in the state, only two were in the pool of schools eligible for TALAS. Either considering the SIG high schools as treated or excluding them from the sample, changes to point estimates are negligible, and the main results are unchanged. We consider these schools untreated in all our presented analyses.

Schools that closed during the study period are included in the treatment samples for the years in which they were open. That is, a school that closed after the 2012-13 school years is included in models for outcomes from 2011-12 and 2012-13, but not 2013-14 or 2014-15. Only six schools identified as lowest-achieving initially followed the closure model, and as a result of their immediate closures they have no outcome data in the years in which this study estimates program effects. An additional six TALAS schools eventually closed during the analysis period[4]—four after 2010-11, two following 2011-12, and two more after 2013-14.

As noted, no schools in North Carolina adopted a restart model, and schools following the closure model have no outcome data for analysis. This study therefore examines the effects of the transformation and turnaround models, which required, among other reforms: leadership change, curricular overhaul, increased recruitment and retention efforts, and coaching-intensive professional development.[5] Specific model-by-year counts for only those schools selected based on proficiency rates are shown in Table 1. Similar model-by-year counts including high schools assigned to turnaround services based on graduation rates are shown in Table 2.

METHODS

Assignment to treatment based on prior proficiency rates creates an opportunity to estimate a local average treatment effect (LATE) at the threshold between treatment and control using a regression discontinuity design (RDD), which exploits local exogeneity in non-random assignment to infer

[4] Local school boards may have considered performance in addition to factors such as funding and enrollment in these closure decisions, but they were not explicitly part of the turnaround intervention.

[5] Although research on the implementation of these models under the SIG program in California finds greater impact from the turnaround model than the transformation model (Dee, 2012), in this report we will not differentiate the two because the turnaround model was used primarily by only one district, confounding the impacts of the model itself with the specific context and capacity of the district. In a separate, co-authored study, we specifically explore the contrast in treatment effects between these two models within this district.

causal impacts from quasi-experimental data (Cook, 2008; Lee & Lemieux, 2009; Van der Klaauw, 2008). In the regression discontinuity framework, an outcome variable (often subsequent measures of the assignment variable) is modeled as a function of the assignment variable. A break, or discontinuity, in this functional relationship at the threshold value of the forcing variable is inferred to be the local treatment effect at the assignment threshold. Notably, this LATE is not an average treatment effect (ATE) and thus cannot be inferred as an overall treatment effect unless treatment effects are assumed to be homogenous. Findings from this study may not generalize to all treated schools, including those lowest-achieving schools furthest from the assignment threshold. The Institute of Education Science's (IES) What Works Clearinghouse (WWC) specifies several conditions which must be met in order for causal inferences to be valid (Schochet et al., 2010). These conditions include integrity of the forcing variable, limited attrition, continuity of the outcome-forcing variable relationship, and the correct identification of the functional form and optimal bandwidth. We describe the extent to which this study meets these standards in the Validity Checks section which appears after the Results section.

We model student achievement in each school in the post-intervention years as a function of the 2009-10 school-level proficiency rate upon which assignment to treatment was determined. The threshold of the forcing variable refers to the value below which schools were assigned to school turnaround, and above which schools were exempt. Because the threshold varied by elementary, middle, and secondary grade levels, we first center the forcing variable within each of these grade levels by subtracting from each school's 2009-10 proficiency rate the threshold value for its grade level. As a result, zero represents the threshold value in each level and the observations across all three school levels can be combined into an equation specified as:

$$(1)\ \ y_{ijkt} = \beta_0 + \beta_1 * x_{it} + \beta_2 * y_{ijkt-1} + \beta_3 * F(PR_{10cj}) + \beta_4 * z_j + \beta_5 * z_j * F(PR_{10cj}) + \varepsilon_{ijt}$$

Where y_{ijkt} represents a standardized achievement score for student *i* in school *j* in subject *k* in year *t*, y_{ikjt-1} is student *i*'s prior year score[6,7] in the same subject, *k*, in year *t-1*, and x_{it} is a vector of student characteristics in year *t*. $F(PR_{10cj})$ is a function of the threshold-centered standardized 2009-10 proficiency rate for school *j* which is optimized to fit the functional form of the data for each model, and *z* is the treatment indicator equal to 1 for students in schools receiving turnaround services and 0 for students in schools that did not qualify. The main effect of treatment is estimated by β_4, the coefficient on the treatment indicator *z*, and captures the discontinuity in the linear relationship between the forcing variable PR_{10c} and the outcome *y* at the threshold. The interaction term between the treatment indicator and the forcing variable allows the slope (relationship between pre- and post-treatment outcomes) represented by β_5 to vary between treatment and control in addition to the discontinuity main effect represented by β_4. Because prior year scores are used as controls in the model, students in treatment schools in both *t* and *t-1* have post-treatment outcomes as baseline controls, meaning any results identified in a given year are above and beyond treatment effects from the prior year. This may attenuate estimated treatment effects, with attenuation increasing in later years as student baseline controls include up to four years of prior treatment effects. Error terms are expressed as ε_{ijt}, without the *k* subject indicators because errors are clustered at the student level for all tests to account for serial correlation of errors across subjects for individual students.

[6] For all EOG tests in fourth through eighth grade reading and mathematics, same-subject scores from the prior year are used as controls. For high school End-of-Course exams, prior scores are taken from students' eighth-grade EOG tests. In some cases, this is a two-year lag.

[7] Including prior scores as a control variable case-wise deletes all observations without such prior scores. To the extent that students without prior test scores, such as out-of-state movers or migrants, are over-represented in treatment schools, this has the potential to bias estimates. As a robustness check, we ran alternative models in which prior scores are interacted with a binary indicator for having data. As a result, missing prior data results in an intercept adjustment for predicted achievement. The results were similar in direction, magnitude, and significance levels, suggesting that missing prior test scores in the sample is not a source of significant bias.

Although assignment to TALAS followed strict assignment to treatment based on 2009-10 school proficiency rates, the addition of seven high schools based on graduation rates below the 60 percent threshold (though only constitutes a potential confounder to the effect estimates (See Figure 1). Because all non-compliance to the proficiency-based assignment threshold pertains to secondary schools, for models limited to only elementary and middle school outcomes, we can estimate strict regression discontinuity designs as specified above. For effect estimates specific to secondary schools or including secondary schools, we address non-compliance in two ways. First we estimate a fuzzy RDD, a two-stage least squares (2SLS) model in which we use 2009-10 school proficiency rates to estimate probability of assignment, *P(z)*:

$$(2) \qquad P(z_j) = \alpha_{00} + X_j\alpha_{xj} + \alpha_{PR}F(PR_{10cj}) + v_{jt}$$

In this first stage equation, the probability of assignment for school j is estimated as a function of the school's student body characteristics x, and its 2009-10 standardized and threshold-centered proficiency rate PR_{10cj}.

The second method for addressing the violation of strict treatment assignment is a frontier regression discontinuity design. Secondary schools, which could be assigned to turnaround for either proficiency or graduation rates, present a multiple rating-score scenario which can be addressed using a "frontier" RDD approach described by Reardon and Robinson (2012) and earlier implemented by Dee (2012) to evaluate the effect of SIG awards. In Dee's study, California schools needed to fall below thresholds of both performance and progress in order to be eligible for SIG awards. By using one criterion to define a "frontier space" of eligible schools and the other as the assignment threshold, this created two separate discontinuities. That is, among all schools below the performance threshold, progress acted as the forcing variable; among all schools below the progress threshold, performance acted as the forcing variable. Because the North Carolina context offers “either” rather than “and” criteria, it presents two analogous opportunities to analyze one of the two assignment variables in isolation within the frontier space of schools identified as eligible by the other. Figure 1 illustrates the frontiers created by the two forcing variables for North Carolina high schools, with proficiency rates on the x-axis, graduation rates on the y-axis, and reference lines indicating the assignment threshold for each measure. Using the 2009-10 proficiency rate as the forcing variable, the cutoff between quadrant I and quadrant II represents a sharp assignment as does the cutoff between quadrant I and quadrant III using the 2009-10 graduation rate as the forcing variable. Thus we estimate frontier RDD treatment effects both at the proficiency threshold among high schools with 2009-10 graduation rates above 60% (labeled “Frontier A”), and at the graduation threshold among high schools with 2009-10 proficiency rates above the 58.1% assignment threshold (labeled “Frontier B”). We consider elementary schools and middle schools without a graduation rate as falling within the Frontier A sample space for all-levels effect estimates, but can only estimate the Frontier B models for secondary schools that fall on the 2009-10 graduation rate forcing variable continuum.

Each model is fitted using the highest polynomial term of the forcing variable found to be significant at $p<0.05$ over all observations. These range from a quadratic to a fourth-order term. Though the functional form of the model is determined using all observations, the same models are applied within optimal restricted bandwidths in semi-parametric models. As a result, higher-order terms of the forcing variable are retained in these models than would be supported within the restricted data spaces.

RESULTS

We present semi-parametric estimates using optimal bandwidth restrictions (OB) and parametric estimates fitted with polynomial functions of the forcing variable across full analytic samples without bandwidth restrictions (Full[8]) in Table 2. The first two rows present overall effect estimates for models combing outcome data across elementary, middle, and secondary schools (All). Of the three models, only the first row—the fuzzy RDD—accounts for the treatment of some high schools selected on the basis of their low graduation rates. When fit across all grades, the fuzzy RDD estimates a significant a positive effect ranging from 0.042 to 0.093 in three of the four post-intervention years. However, within optimal bandwidths more narrowly focused around the assignment threshold, the estimates are not statistically significant in two of the outcome years and significant negative in the other two outcome years, with effect estimates of -0.050 in 2013-14 and -0.039 in 2014-15. The frontier RDD estimates significant and modest negative effects in the first three years of outcome data (OB), and significant negative effects effects in 2011-12, 2012-13, and 2014-15 (Full).

In the third row of Table 2, effect estimates for elementary and middle schools are presented (secondary schools omitted). The strict assignment of elementary and middle schools without any non-complying cases supports a sharp RDD, and the effect estimates are significant and negative in each of the four outcome years, ranging from -0.020 to -0.056 for all but the full 2012-13 model.

Rows four through six present the TALAS effect estimates for high schools. In contrast to elementary and middle schools, 17 out of 24 effect estimates are positive and significant, with positive significant effects ranging from 0.034 to 0.198 standard deviation units. Seven of the 17 significant effect estimates are above ten percent of a standard deviation unit. The largest and most consistent effects in secondary schools are found in the fuzzy RD model, which finds significant and

[8] To reduce confusion between models in which elementary, middle, and secondary schools are analyzed together from those in which we do not apply bandwidth restrictions to a given analytic sample defined by grade level, we use “All” to refer to analyses in which all grade levels are combined, and “Full” to refer to models in which no bandwidth restrictions are applied.

positive effects in each of the four outcome years using either optimal bandwidth restrictions (OB) or the full analytic sample without bandwidth restriction (Full). The secondary Frontier A model, which excludes both treated and untreated low-graduation rate high schools, also finds treatment gains, with significant and positive effects in both the optimal bandwidth and full sample models in 2011-12 and 2014-15 ranging from 0.057 to 0.111, positive effects in 2012-13 that are only significant over the full sample, but negative and significant effects in both OB and Full in 2013-14.

The Frontier B model, in which treatment schools were selected by 2009-10 graduation rates, finds positive effects in seven of the eight models, with three of the four in the first two years of outcome data statistically significant. However, it also finds a significant negative effect (-0.238) within the bandwidth-restricted model. The smaller sample size in this model leads to less precise estimates, especially within restricted bandwidths. Notably, the two models that find the most consistent positive effects—the fuzzy RD and the Frontier B—account for the treatment status of high schools falling above the proficiency threshold but assigned to treatment by graduation rates.

Checks on the Validity of the RDD

The What Works Clearinghouse sets five standards against which the validity of regression discontinuity studies can be examined for validity. Using these standards as a guide, we find that the design and implementation of TALAS support the use of RDD, and that the positive effect estimates in the secondary models are robust to a number of statistical tests. The first standard addresses the assumption of conditionally exogenous treatment at the assignment threshold by testing the integrity of the forcing variable. A number of contextual factors support this assumption and satisfy the first standard. First, the assignment process, as described in the Background section above, indicates that the schools could not manipulate their treatment status since the criterion for assignment was not determined prior to the testing that created the running variable. Instead, the federal government set the cut-off and it was thus not subject to the potential for local manipulation, and schools would not have had knowledge of where their scores would fall with respect to the cut-off before the testing. This is also supported by the density plot of schools' values of the forcing variable in Figure 2, which shows no unusual clumping of schools on either side of the threshold. AMcCrary density test (McCrary, 2008) finds an insignificant log difference in the height of the distribution function at the assignment threshold (log diff =0.20, se=0.23), failing to reject the null hypothesis of non-manipulation (p=0.96).

The second standard for RDD studies is that attrition rates should not be significantly different between treatment and control groups within optimal bandwidths of the forcing variable. The inclusion of "closure" as a treatment model complicates the compliance of this standard, as the turnaround program gave the state the option to close a failing school and reassign its students as a form of turnaround. As a result of this feature of the turnaround program, the attrition rate among schools initially designated for turnaround is roughly twice that of all other schools in the state during the treatment period. Table 4 shows differences in the attrition rates between treated and non-treated schools by year and bandwidth. The optimal bandwidth of the proficiency rate forcing variable varies by year, analysis sample, and model specification, but is generally closest to 1.0 sdu among the critical values shown in Table 4. Unfortunately, the difference in attrition is significant at standard thresholds even within this restricted range. While the study does not satisfy this standard, we argue that the deliberate closure of schools as part of the treatment is not the concern the standard was written to address.

With adjustments to the threshold across the *x*-axis, we only find statistically significant discontinuity near an adjustment of zero, or when the assignment threshold matches its actual value. We also test the sensitivity of the estimate to changes in the bandwidth of the forcing variable within which we estimate the model. As Figure 5 shows, the estimate is positive and significant within any bandwidth, though it becomes much less precise when restricted to less than one standard deviation of the forcing variable on either side of the assignment threshold. This loss of precision results from the smaller sample sizes included within more restrictive bandwidths, but the consistent positive effect estimates assure that our findings are not sensitive to sample exclusions. A complete list of the What Works Clearinghouse standards and their relation to TALAS is presented in Table 4.

In Figure 3, we show that the relationship between the forcing variable(the proficiency rate in 2009-10) and the proficiency rate in 2008-09 is smooth at the cutoff as a way of simulating what the relationship between the forcing variable and the outcome variable would be in the absence of treatment.

DISCUSSION

The DST's implementation of its TALAS initiative presents an important opportunity to evaluate the potential for state education agencies to implement current federally-approved turnaround models to improve academic outcomes in chronically low-achieving schools. Such initiatives have gained greater priority status in federal education reform efforts over the past decades, with policymakers still eager to find evidence of effective strategies. Where other initiatives have lacked ambitious reforms, implementation fidelity, stakeholder commitment, or agency support, North Carolina's previously-established division for school turnaround, stakeholder commitment, and implementation fidelity to comprehensive turnaround efforts present a more ideal situation in which to test the potential efficacy of comprehensive school turnaround efforts.

Overall, results are more mixed across grade levels than across turnaround models. We find consistent, positive effects on student achievement in secondary schools sustained in each of the four years in the post-turnaround period. Conversely, we find modest negative effects for students in elementary and middle schools served by TALAS. When including all grade levels, the effects vary by bandwidth. The analysis indicates that the wider the bandwidth, the more positive the effects, perhaps suggesting that effects were heterogeneous, with larger effects occurring in the lowest performing schools.

Several aspects of the setting for this study may account for the more positive effects among high schools. First and as previously mentioned, North Carolina's lowest performing high schools had been the focus of a previous round of turnaround ordered by the court. While some middle schools and elementary schools were also in turnaround under the predecessor to TALAS, that intervention was designed specifically for high schools, and high schools received more turnaround services over a longer period. For example, the original turnaround framework included freshman academies, which are only applicable to high schools. It may be that the results for TALAS were greater for high schools due to the DST's experience working with and the focusing on high schools. On the other hand, the fact that the high schools had the highest absolute cutoff on the forcing variable indicates that they had lower levels of non-proficiency than the middle and elementary schools served by TALAS.

We note three other issues that readers should factor into any conclusions from this study or others on school turnaround. The first pertains to understanding the nature of the intervention, the second relates to the statistical method used to evaluate its effectiveness, and the third considers how we interpret the outcome measure(s) selected. First, on the nature of the intervention, Dee (2012) found that in California, the intervention model with stricter requirements, "turnaround," was effective while the transformation model was not. In North Carolina as in most other RttT states, the transformation model was most prevalent, yet may be too weak to produce consistent effects. Other studies have looked at a single turnaround model or have not examined differential facets of the interventions studied. Both Strunk, et al. (2016) and Zimmer, et al. (2016) suggest that a school climate supporting innovation and continuous improvement can produce effects, but sustaining those effects may depend on the ability to sustain that climate and build human and social capital.

The second issue in interpreting findings on school turnaround relates to the estimation method used by the researchers. When assumptions are met, RDD provide credible causal effect estimates that are local average treatment effects (LATE) at the assignment threshold. Because these LATE estimates do not necessarily generalize to all treated observations as an average treatment effect (ATE), they may be better interpreted as an evaluation of whether the assignment threshold was set at an optimal value than as an evaluation of the intervention's effectiveness *per se.* In the present study, a plausible interpretation of the positive effects on high schools is that the cutoff for high schools was set such that students in schools just missing assignment would have been better off in a school assigned to turnaround. That is, schools above the assignment threshold may also have benefited from TALAS, and thus an optimal assignment threshold would have included more schools. Similarly, the negative effects on elementary and middle school achievement may not reflect detrimental or even ineffective turnaround efforts, but that the cutoff for elementary and middle schools (52.3 and 53.3 percent proficient, respectively) may have been too high, and an optimal assignment threshold would have included fewer schools. Additional research that can explore effect heterogeneity at distance away from the cutoff may be needed to better understand the effects on the lowest performing of the schools in turnaround. These studies will not have the strong claims to internal validity that RDD achieves, but they may be useful to suggest hypotheses about the schools that stand to benefit the most from state turnaround efforts.

Finally, summative evaluations of school turnaround using only student achievement scores on state assessments are limited at best. While achievement scores offer an important summative measure of students' learning, they also have multiple shortcomings. Given the range of tested grades and subjects, models using state achievement scores cannot account for gains in grades K-3, most of 11th and 12th grades, or any non-tested subjects. Collectively, these gaps constitute the majority of students' K-12 educational careers. Further, estimates of achievement gains measure academic performance on a different scale and by a different metric than the one states typically use to designate schools as low-performing. As such, these evaluation measure progress in units (student-level achievement gains) that cannot speak to the initial conceptualization of the problem (school-level proficiency rates). Even large positive achievement gains may not result in meaningful changes in these schools' relative standing in the distribution of statewide proficiency rates. Without reconciling the tension between states defining schools as low-performing by one measure and assessing their progress by another, the tendency for even those turnaround schools demonstrating statistically significant achievement gains to remain near the bottom of statewide proficiency distributions may perplex all stakeholders, including evaluators, policymakers, educators, and parents.

Future work should continue to improve measurement of outcomes with the potential to reflect more potential benefits of turnaround efforts and to create study designs capable of offering not only strong causal warrant, but also results of greatest relevance to educators, policymakers, and K-12 learners. Without random assignment of schools to turnaround, the most credible and influential estimates of state turnaround will continue to be illusive.

References

Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis. Which high schools produce the nation's dropouts? Where are they located? Who attends them? Report 70*. Baltimore, M.D.: Johns Hopkins University: Center for Research on the Education of Students Placed At Risk, Publications Department, CRESPAR. Retrieved from http://eric.ed.gov/?id=ED484525

Berends, M., Bodilly, S., & Kirby, S. N. (2002). Looking back over a decade of whole-school reform: The experience of New American Schools. *The Phi Delta Kappan*, *84*(2), 168–175.

Bifulco, R., Duncombe, W., & Yinger, J. (2005). Does whole-school reform boost student performance? The case of New York City. *Journal of Policy Analysis and Management*, *24*(1), 47–72.

Brownstein, R. (2003). Locked Down. *Education Next*, *3*(3), 40–46.

Cook, T. D. (2008). "Waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, *142*(2), 636–654.

Dee, T. (2012). School turnarounds: Evidence from the 2009 stimulus. *National Bureau of Economic Research*, (Working Paper 17990).

Duncan, A. (2009, June 22). Turning Around the Bottom Five Percent [Speeches and Testimony]. Retrieved October 22, 2014, from http://www2.ed.gov/news/speeches/2009/06/06222009.html

Gross, B., Booker, T. K., & Goldhaber, D. (2009). Boosting student achievement: The effect of comprehensive school reform on student achievement. *Educational Evaluation and Policy Analysis*, *31*(2), 111–126.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). Turning Around Chronically Low-Performing Schools. IES Practice Guide. NCEE 2008-

4020. *National Center for Education Evaluation and Regional Assistance*. Retrieved from http://eric.ed.gov/?id=ED501241

Imbens, G., & Kalyanaraman, K. (2011). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, rdr043. http://doi.org/10.1093/restud/rdr043

Lee, D. S., & Lemieux, T. (2009). *Regression discontinuity designs in economics*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14723

Mintrop, H., & MacLellan, A. M. (2002). School Improvement Plans in Elementary and Middle Schools on Probation. *The Elementary School Journal*, *102*(4), 275–300.

Mintrop, H., MacLellan, A. M., & Quintero, M. F. (2001). School Improvement Plans in Schools on Probation: A Comparative Content Analysis across Three Accountability Systems. *Educational Administration Quarterly*, *37*(2), 197–218. http://doi.org/10.1177/00131610121969299

Murphy, J., & Meyers, C. V. (2008). *Turning Around Failing Schools: Leadership Lessons From the Organizational Sciences*. Thousand Oaks, C.A.: SAGE Publications.

North Carolina Department of Public Instruction. (2010). *Computing the ABCs Performance Composite 2009 10*. Raleigh, NC: North Carolina Department of Public Instruction. Retrieved from http://www.ncpublicschools.org/docs/accountability/reporting/abc/2009-10/composite.pdf

Papay, J. (2015). The Effects of School Turnaround Strategies in Massachusetts. In *2015 Fall Conference: The Golden Age of Evidence-Based Policy*. Appam. Retrieved from https://appam.confex.com/appam/2015/webprogram/Paper14077.html

Reardon, S. F., & Robinson, J. P. (2012). Regression Discontinuity Designs With Multiple Rating-Score Variables. *Journal of Research on Educational Effectiveness*, *5*(1), 83–104. http://doi.org/10.1080/19345747.2011.609583

Rothstein, R. (2013). *For public schools, segregation then, segregation since: Education and the unfinished march*. Economic Policy Institute. Retrieved from http://www.epi.org/publication/unfinished-march-public-school-segregation/

Schochet, P. Z., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for Regression Discontinuity Designs. Retrieved January 13, 2015, from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=231

Schueler, B. E., Goodman, J., & Deming, D. J. (2016). *Can States Take Over and Turn Around School Districts? Evidence from Lawrence, Massachusetts*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w21895

Strunk, K. O., Marsh, J. A., Hashim, A. K., & Bush-Mecenas, S. (2016). Innovation and a Return to the Status Quo A Mixed-Methods Study of School Reconstitution. *Educational Evaluation and Policy Analysis*, 162373716642517.

Thompson, C. L., Brown, K. M., Townsend, L. W., Henry, G. T., & Fortner, C. K. (2011). *Turning around North Carolina's lowest achieving schools (2006-2010)*. Raleigh, NC: Consortium for Educational Research and Evaluation-North Carolina.

Thompson, C. L., Henry, G. T., & Preston, C. (2016). School turnaround through scaffolded craftsmanship. *Teachers College Record (Yearbook)*, *118*(4).

U.S. Department of Education. (2001, January). School improvement report: Executive order on actions for turning around low-performing schools. Washington, DC: Office of Elementary and Secondary Education.

U.S. Department of Education. (2008). Evaluation of the Comprehensive School Reform Program Implementation and Outcomes Third-Year Report. Washington, DC: Office of Planning, Evaluation and Policy Development.

US Department of Education. (2011). SIG-Awarded Schools (SY 10-11, 11-12, 12-13, 13-14). Retrieved from http://www2.ed.gov/programs/sif/index.html

US Department of Education. (2014). *Race to the Top Annual Performance Report*. Retrieved from https://www.rtt-apr.us/

Van der Klaauw, W. (2008). Regression–discontinuity analysis: a survey of recent developments in economics. *Labour*, *22*(2), 219–245.

Zimmer, R., Henry, G. T., & Kho, A. (2016). The role of governance in school turnaround policies: The case of Tennessee's Achievement School District. Presented at the Association for Education Finance and Policy Annual Conference, Denver, CO.

Figure 1: TALAS assignment by 2009-10 proficiency rate (forcing variable).

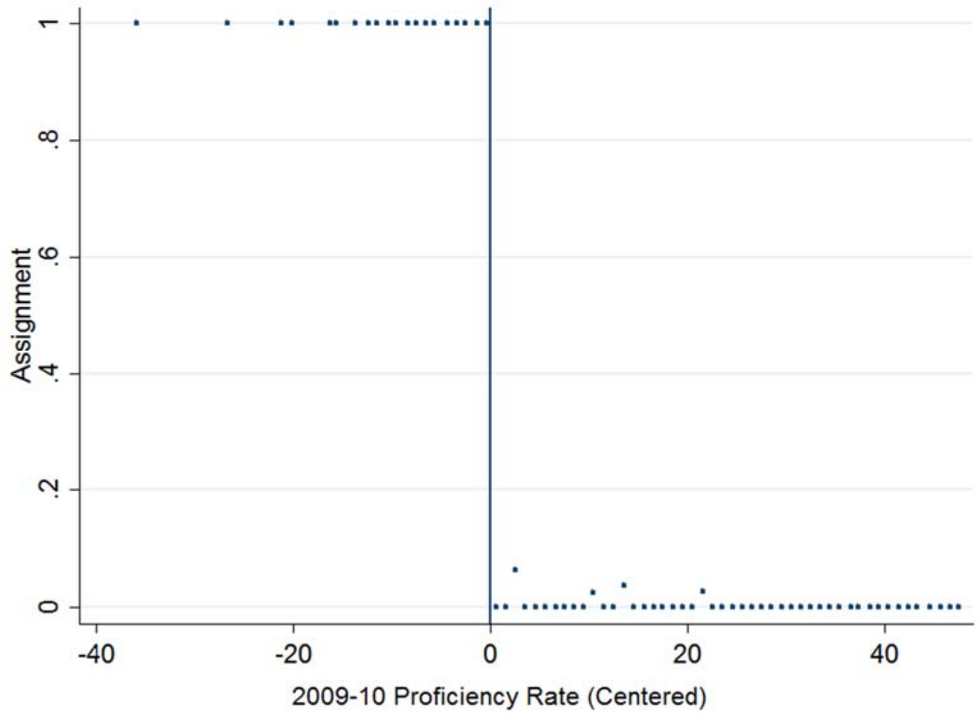

Figure 2. Treatment high schools by assignment variable (and Cartesian plane quadrant numerals).

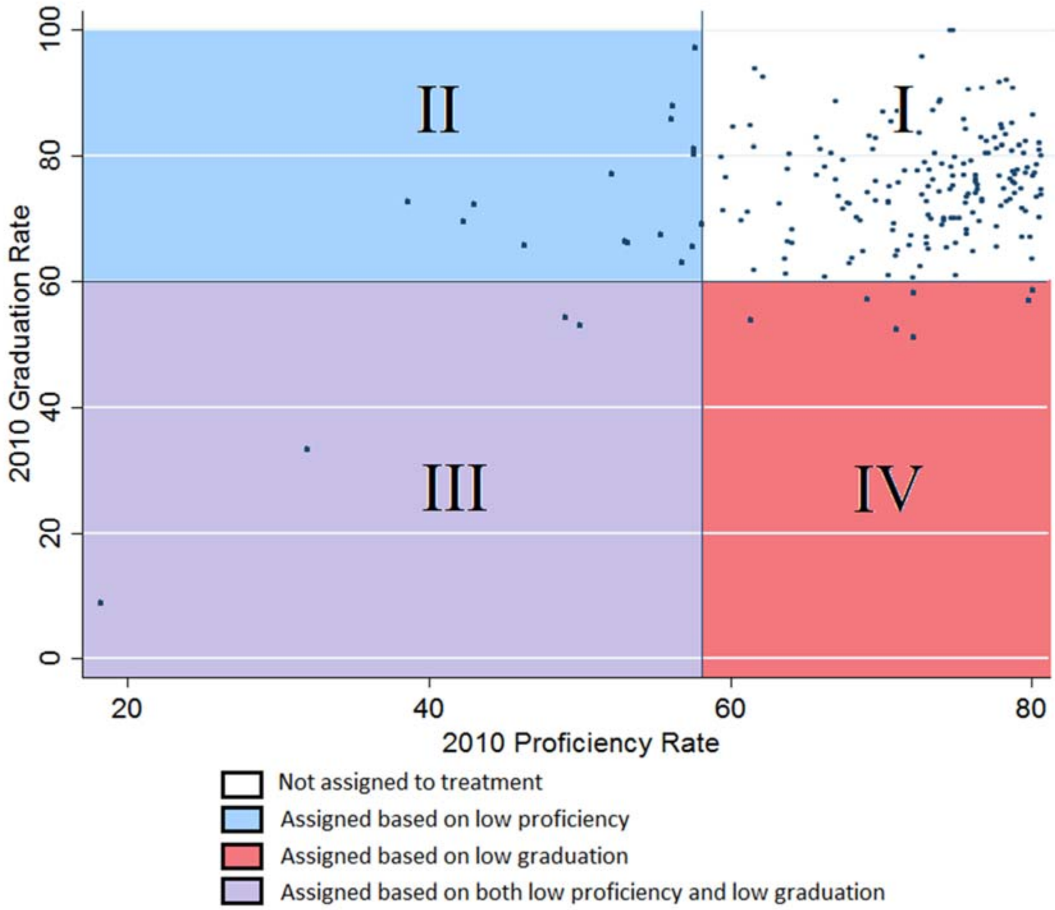

Figure 3. Histogram of 2009-2010 passing rates for schools eligible for assignment, centered at assignment threshold.

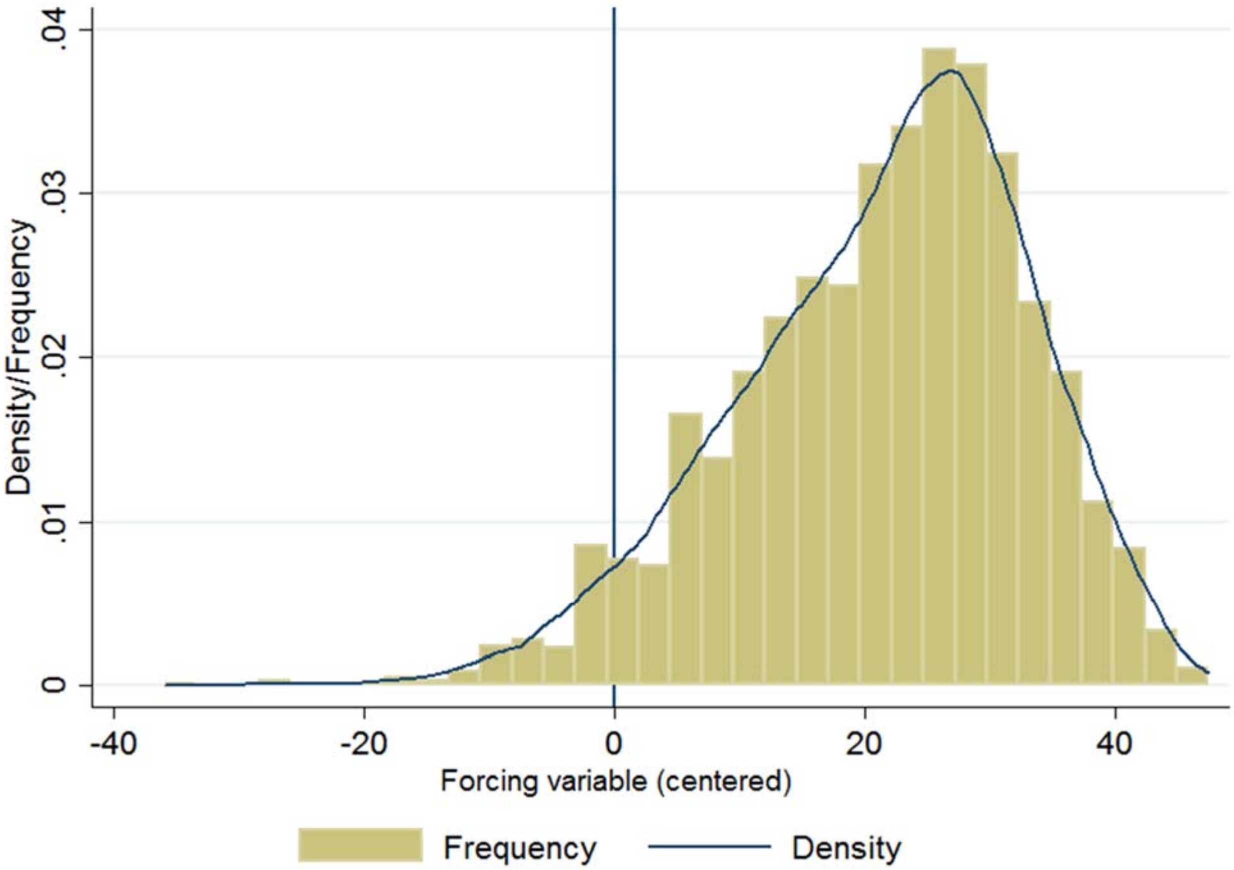

Figure 4. Relationship between 2008-2009 and 2009-10 passing rates (placebo).

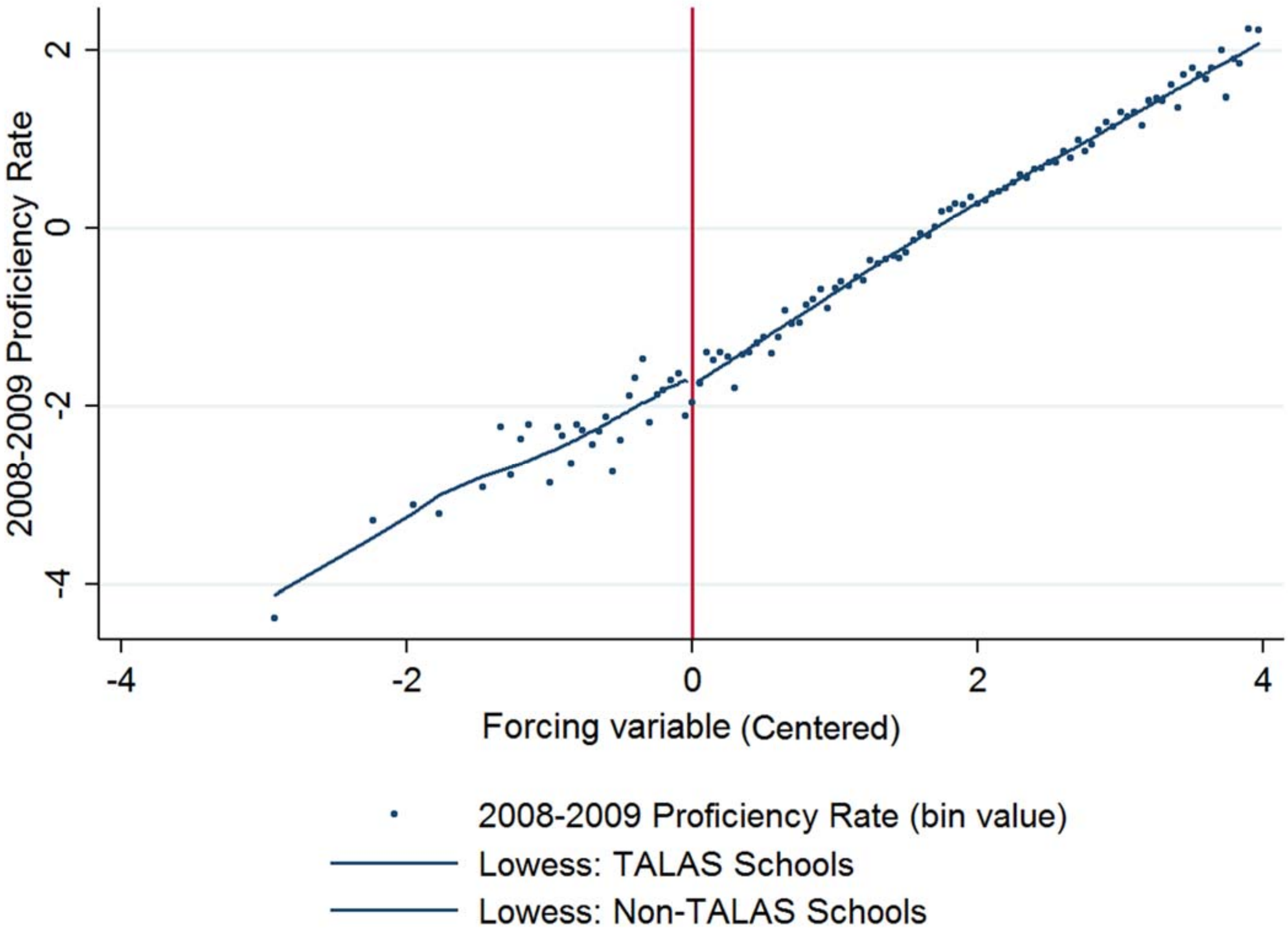

Figure 5. Effect Size by cut-score adjustment: 2014-2015 High schools, frontier RDD model

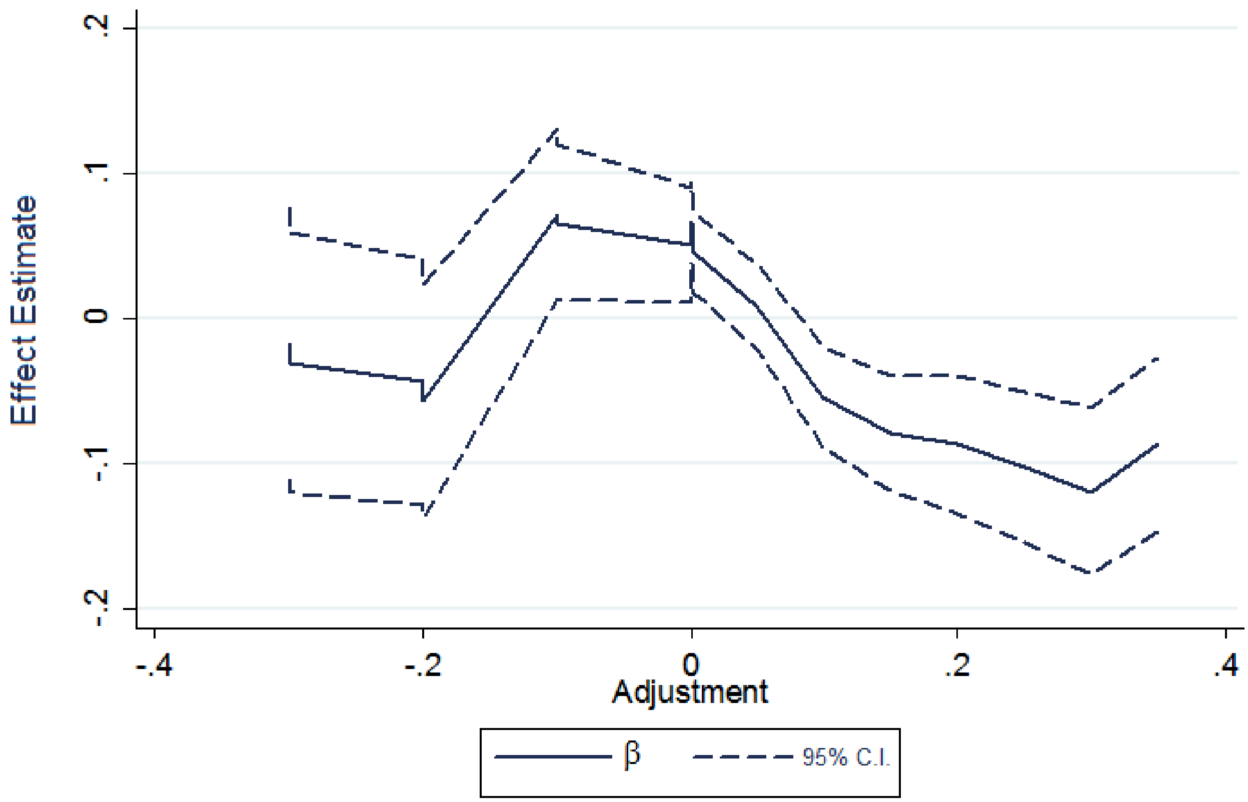

Figure 6. Effect size by bandwidth: 2014-15 high schools, Frontier A model

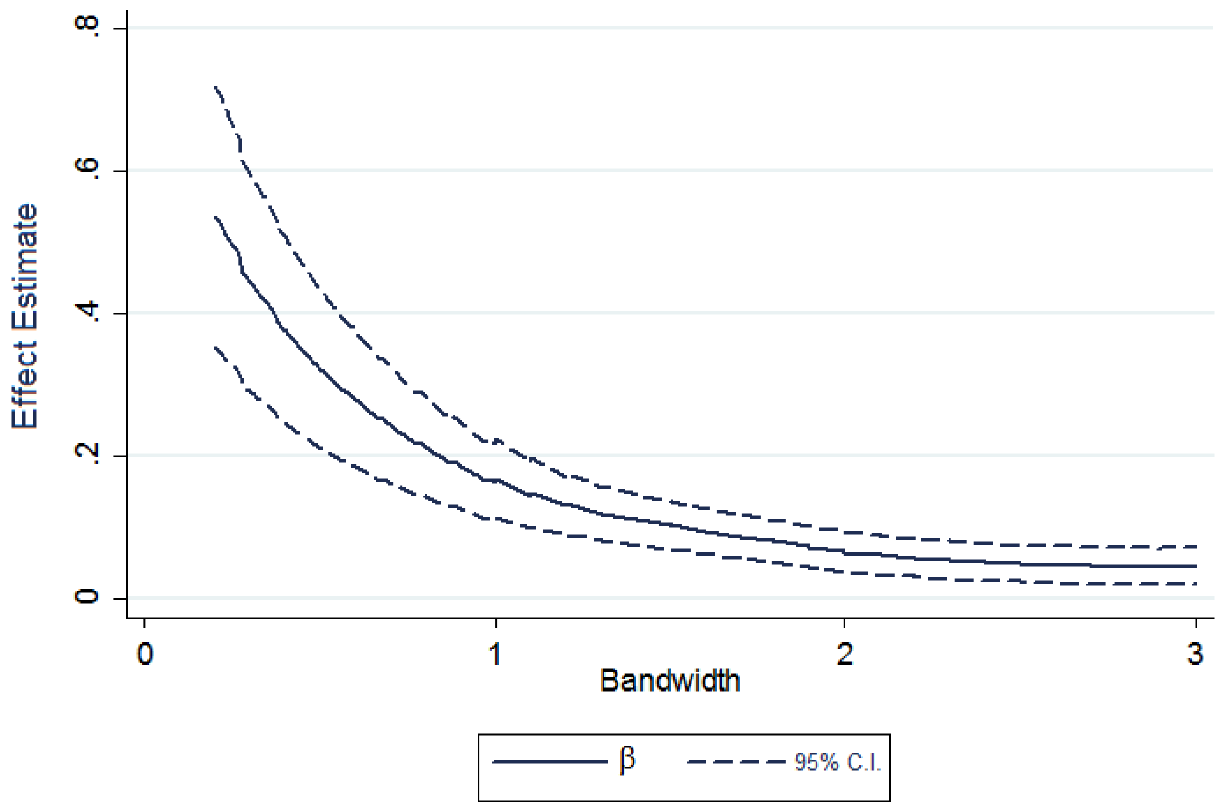

Figure 7. Plotted discontinuity: 2014-15 test scores by bin, all North Carolina schools

Table 1: North Carolina TALAS school turnaround models by year (2010-2011 through 2014-2015)[1]

	2010-11	2011-12	2012-13	2013-14	2014-15
Transformation	102	97	95	93	93
Turnaround	9	10	11	11	11
Closure[2]	6	4	2	2	0
Restart[3]	1	1	0	0	0
TOTAL	118	112	108	106	104

[1] Schools identified based on 2009-10 proficiency rates and graduation rates.
[2]Schools assigned for closure are not counted in years following closure.
[3] One school listed as designated for Restart on North Carolina's Accountability Progress Reporting website in 2010 and 2011. This school never underwent restart and was re-categorized as Transformation in 2012.

Table 2: Summary table of effect estimates on student achievement gains by year, model, and sample.

	2011-12		2012-13		2013-14		2014-15	
	OB	Full	OB	Full	OB	Full	OB	Full
	β/(se)	β/(se)	β/(se)	β/(se)	β/(se)	β/(se)	β/(se)	β /(se)
All Grades (Fuzzy)	-0.012	-0.010	-0.003	0.042 ^{3}	-0.050 ^{3}	0.064 ^{3}	-0.039 ^{1}	0.093 ^{1}
	(0.014)	(0.013)	(0.010)	(0.012)	(0.014)	(0.013)	(0.015)	(0.013)
All Grades (Frontier A)	-0.021 ^{2}	-0.022 ^{3}	-0.021 ^{2}	-0.002	-0.047 ^{3}	-0.019 ^{2}	-0.014	-0.011 ^{1}
	(0.010)	(0.006)	(0.008)	(0.006)	(0.009)	(0.006)	(0.009)	(0.005)
Elementary & Middle	-0.038 ^{3}	-0.028 ^{3}	-0.029 ^{2}	-0.010	-0.056 ^{3}	-0.023 ^{3}	-0.020 ^{1}	-0.028 ^{3}
	(0.010)	(0.006)	(0.009)	(0.006)	(0.010)	(0.006)	(0.010)	(0.005)
Secondary (Fuzzy)	0.055 ^{3}	0.100 ^{3}	0.081 ^{3}	0.174 ^{3}	0.052 ^{2}	0.198 ^{3}	0.034 ^{1}	0.123 ^{3}
	(0.014)	(0.012)	(0.014)	(0.011)	(0.017)	(0.011)	(0.014)	(0.009)
Secondary (Frontier A)	0.083 ^{2}	0.057 ^{2}	0.008	0.043 ^{1}	-0.108 ^{3}	-0.036 ^{1}	0.111 ^{3}	0.064 ^{3}
	(0.029)	(0.019)	(0.022)	(0.017)	(0.026)	(0.017)	(0.022)	(0.015)
Secondary (Frontier B)	0.033	0.140 ^{3}	0.086 ^{1}	0.110 ^{3}	0.041	0.031	-0.238 ^{3}	0.004
	(0.060)	(0.025)	(0.034)	(0.023)	(0.032)	(0.022)	(0.047)	(0.019)

^{1} $p<0.05$, ^{2} $p<0.01$, ^{3} $p<0.001$

Table 3: What Works Clearinghouse Regression Discontinuity Standards Compliance

RDD CRITERIA	COMPLIANCE SUMMARY
Treatment assignments based on a forcing variable	Schools selected on basis of 2009-2010 passing rate. Additional seven high schools selected based on 2009-2010 graduation rates
Forcing variable ordinal with sufficient number of unique values	Passing rates and graduation rates provide unique values of a ratio measure, which supersedes ordinal measurement.
No factor confounded with the forcing variable	SIG grants were only other state intervention specific to low-performing schools, but went to alternative schools not in sample.

STANDARD	COMPLIANCE SUMMARY
1. Integrity of Forcing Variable	
A. Institutional Integrity	* 5% norm-referenced forcing variable harder to manipulate than criterion-referenced * Schools not aware of threshold prior to testing * Threshold specified by ED, not NCDPI * Schools already had incentive to maximize forcing variable
B. Statistical Integrity	McCrary density test shows no discontinuity of the density of schools on either side of the threshold, supporting no manipulation of treatment status
2. Attrition	Attrition rates not significant different between treatment and control within optimal bandwidths for the RD
3. Continuous relationship between forcing variable and outcome in absence of treatment	* Continuous relationship in years prior to treatment * No discontinuity in relationship between forcing variable and key school demographic variables * Results consistent with "placebo" cutoffs, meaning no unexplained discontinuities at other points on the forcing variable
4. Specification	
A. Functional form	* Highest-order polynomial terms of the forcing variable found to be significant in full sample are applied within restricted bandwidths, overfitting model * Highest-order term for model and outcome across years applied for all years.
B. Optimal bandwidth	* Imbens & Kalyanaraman (2011) method for optimal bandwidth identification applied * Where insufficient observations for IK method, models use smallest bandwidth that included all treatment observations

Table 4. Differential attrition by bandwidth of forcing variable around assignment threshold.

	FULL	bw=2.0	bw=1.5	bw=1.0	bw=0.5
2010-11	0.004	0.006	0.012	0.017	0.028
2011-12	-0.063 [3]	-0.067 [3]	-0.059 [1]	-0.051 [1]	-0.016
2012-13	-0.067 [3]	-0.069 [2]	-0.061 [2]	-0.049 [1]	-0.016
2013-14	-0.100 [3]	-0.103 [3]	-0.087 [3]	-0.076 [2]	-0.052
2014-15	-0.107 [3]	-0.109 [3]	-0.092 [3]	-0.080 [2]	-0.060

[1] $p<0.05$, [2] $p<0.01$, [3] $p<0.001$

Differences calculated by subtracting the attrition rate (schools not present in a given year that were present in the 2009-2010 baseline year) among TALAS schools from the attrition of non-TALAS schools.
Results are presented as differences in proportions.
Bandwidths stated in standard deviation units.

Table 5. Discontinuity of forcing variable-covariate relationship at assignment threshold.

Covariate (2011-2012)	β	S.E.
Per-pupil expenditure	0.036	(0.03)
Free or reduced-price lunch	0.089	(0.11)
Minority	0.051	(0.60)
Average daily membership	-0.227	(0.14)

Effects and standard errors are in standard deviation units
[1] $p<0.05$, [2] $p<0.01$, [3] $p<0.001$

Table 6. Highest significant polynomial order of forcing variable by sample, model, and year.

	2011-12	2012-13	2013-14	2014-15	Highest
All Grades (Fuzzy)	2nd	3rd	3rd	3rd	3rd
All Grades (Frontier A)	2nd	3rd	3rd	3rd	3rd
Elementary & Middle	3rd	3rd	3rd	3rd	3rd
Secondary (Fuzzy)	5th	3rd	3rd	3rd	5th
Secondary (Frontier A)	4th	3rd	3rd	3rd	3rd
Secondary (Frontier B)	2nd	2nd	2nd	2nd	2nd

Table 7. Optimal bandwidth of forcing variable by sample, model, and year.

	2011-12	2012-13	2013-14	2014-15
All Grades (Fuzzy)	1.166	1.565	1.197	0.958
N (Treatment/Total)	(95/563)	(98/835)	(95/575)	(85/420)
All Grades (Frontier A)	1.089	1.274	1.123	0.977
N (Treatment/Total)	(89/503)	(90/624)	(88/519)	(83/425)
Elementary & Middle	1.119	1.201	1.074	0.902
N (Treatment/Total)	(79/430)	(78/467)	(78/401)	(72/331)
Secondary (Fuzzy)	1.551	1.324	1.078	1.057
N (Treatment/Total)	(18/175)	(15/127)	(14/83)	(14/81)
Secondary (Frontier A)	1.036	1.443	1.028	1.066
N (Treatment/Total)	(10/77)	(11/146)	(10/73)	(10/77)
Secondary (Frontier B)	0.389	0.443	0.457	0.246
N (Treatment/Total)	(4/24)	(5/28)	(5/29)	(4/15)