THE NATIONAL CENTER ON
SCALING UP
EFFECTIVE SCHOOLS

Can High Schools Be Turned Around? Impacts of State Led Turnaround on Student Achievement, Graduation and Absenteeism

Gary T. Henry | Roderick A. Rose | Shanyce L. Campbell

Conference Paper

June 2012

Prepared for Achieving Success at Scale: Research on Effective High Schools in Nashville, Tennessee

VANDERBILT PEABODY COLLEGE

EDC Learning transforms lives.

Florida State University 1851

Georgia State University

The University of North Carolina at Chapel Hill

WISCONSIN University of Wisconsin–Madison

The National Center on Scaling Up Effective Schools (NCSU) is a national research and development center that focuses on identifying the combination of essential components and the programs, practices, processes and policies that make some high schools in large urban districts particularly effective with low income students, minority students, and English language learners. The Center's goal is to develop, implement, and test new processes that other districts will be able to use to scale up effective practices within the context of their own goals and unique circumstances. Led by Vanderbilt University's Peabody College, our partners include The University of North Carolina at Chapel Hill, Florida State University, the University of Wisconsin-Madison, Georgia State University, and the Education Development Center.

This paper was presented at NCSU's first national conference, *Achieving Success at Scale: Research on Effective High Schools*. The conference was held on June 10-12, 2012 in Nashville, TN. The authors are:

Gary T. Henry
Roderick A. Rose
Shanyce L. Campbell
University of North Carolina at Chapel Hill

This research was conducted with funding from the Institute of Education Sciences (R305C10023). The opinions expressed in this article are those of the authors and do not necessarily represent the views of the sponsor or the National Center on Scaling Up Effective Schools.

Can High Schools Be Turned Around? Impacts of State Led Turnaround on Student Achievement, Graduation and Absenteeism

Abstract

In this study, we present rigorous evidence of the impact of efforts to turn around high schools on a scale that could potentially reduce the number of very low performing schools. In 2006-06 and 2007-08, the North Carolina Department of Public Instruction initiated efforts to turn around 66 chronically low performing high schools (14% of the regular high schools in North Carolina) as a result of judicial and accountability pressures. We use a fuzzy regression discontinuity design with a difference-in-differences estimation framework that allows us to leverage a panel data set that includes at least two years prior to the intervention and three years after to estimate the effects of turnaround. The school turnaround intervention improved on the state's performance composite rating and test scores in the five core courses required for students to graduate—Algebra1, English 1, Biology, Civics and U.S. History – but the effects on student absenteeism and graduation rates were not significant. The effects on test scores were very heterogeneous and more research is needed to identify any systematic sources of the heterogeneity.

Can High Schools Be Turned Around? Impacts of State Led Turnaround on Student Achievement, Graduation and Absenteeism

Perhaps no one has more powerfully or concisely expressed the significance of efforts to turnaround America's failing schools than Nicholas Lemann, "We have a moral obligation to be precise about what the problems in American education are – like subpar schools for poor children and minority children – and to resist heroic ideas about what would solve them, if the ideas don't demonstrably do that" (2010, np). Lemann points out the two sides of the challenge for public education, correctly identifying a problem and a solution that actually addresses the problem at a sufficient scale to matter. Evidence shows that a compelling problem confronting education in the U.S. is that many schools that serve high concentrations of poor and racial and ethnic minority students are highly ineffective, which makes identifying and intervening in these schools a logical approach to one of the main problems facing education in the U.S. One effort to identify these schools has focused on schools with low graduation rates, which Balfanz and Legters label "dropout factories" (2004). While some progress was made between 2002 and 2008 in improving the "promoting power" in dropout factories, the number of schools with graduation rates below 60 percent has stubbornly remained constant (Balfanz & Legters, 2004; Balfanz, Bridgeland, Moore & Fox, 2010). Based on accountability pressures at the federal and state level, another focus is schools with low standardized test scores. Although there is a high correlation between schools with low graduation rates and low test scores, schools are usually more narrowly classified as "low performing" based on failure to meet standardized test score targets.

In Lemann's statement there is also the call to resist attractive solutions that promise too much and deliver too little – either in terms of actually producing positive effects or scale of operations. He includes among these attractive solutions those that are highly effective but unlikely to reach a sufficient scale to address the problem such as the Harlem Children's Zone and Teach For America and those that resonate with policymakers and the public such as charter schools that have not been shown to be more effective than the public schools they replace (2010). One solution that seems logical is to intervene directly and dramatically to turn around the chronically low performing schools that are failing large numbers of poor and racial and ethnic minority students. However, this potential solution has yet to be sufficiently tested to determine if it can improve students' educational outcomes on a scale that actually begins to reduce the number of dropout factories and otherwise low performing schools or turns out to be another heroic solution that lacks results.

Turning around chronically low performing districts and schools is an important yet arduous task. The varied typologies of federal and local approaches that fall under the turnaround label further complicate our understanding of the most effective way to turn around schools or even having a definition of the turnaround approach. Because research on turnaround schools is currently limited in quantity and rigor, it remains unclear what definition of school turnaround can be used for research purposes and whether the turnaround approach effectively improves student achievement (Herman et al. 2008, Murphy, 2009).

Turnaround is often characterized by replacement of school leadership and large proportions of teachers, a focus on improving instruction and achieving visible improvements rapidly (Herman et al. 2008). These are policy actions and descriptions of the unfolding process rather than a concrete series of actions that, in and of themselves, would lead directly to improved student performance. Turnaround is not an intervention in the sense that a curriculum change is an intervention. It is defined by the original status of the school and the rapid change in that status. This is fine for the retrospective research that characterizes case studies where the definition is used to identify schools that turned around in order to identify the possible mechanisms and generate hypotheses about what led to the turnaround. But for prospective research, we define turnaround schools as schools placed in a turnaround program based on low levels of student performance with the expectation for rapid improvement. We cannot use rapid improvement as both a part of the definition **and** the outcome by which turnaround is evaluated. Similar to Herman et al. (2008), this study does not address Comprehensive School Reforms (CSR) – school-wide models for reform previously supported by the federal government and designed to improve student achievement. The research on CSR has shown some promising results, however these findings are based on models aimed at school improvement that generally do not include or at least seldom mandate personnel replacement as a key component (Borman, Hewes, Overman, & Brown, 2003). Consistent with the literature on school turnaround to date (Herman 2008), we take an operational definition of school turnaround to be schools with a high proportion of low performing students who rapidly decrease the proportion of low performing students.

In addition, much of existing literature on school turnaround is based on case studies, which do not allow for cause-effect conclusions to be drawn about the turnaround intervention. Additionally, the results from these case studies may not be generalizable in the larger context. This study advances the literature by causally examining the effectiveness of the turnaround model. Before scholars began to tease apart which components of turnaround models are more effective than others, it is important that determine whether the turnaround intervention is effective in improving student achievement. In this study, we present rigorous evidence of the impact of efforts to turn around North Carolina high schools on a scale that could potentially reduce the number of very low performing schools. In North Carolina, schools with very low performance composite scores (hereafter "PC") are classified as "low performing." The PC is the number of End-of-Course tests passed as a percentage of those taken. In 2006-07 and 2007-08, the North Carolina Department of Public Instruction (NC DPI) initiated efforts to turn around 66 chronically low performing high schools (14% of the regular high schools in North Carolina) as a result of judicial and accountability pressures. Because the low performing high schools were assigned to the state's turnaround intervention based on a measure of their students' academic performance, the study employed a regression discontinuity design to estimate the causal effect of the intervention.

We leveraged a longitudinal, student-level dataset that included a minimum of two years of pre-intervention and a minimum of three years of post-intervention data to estimate the impacts strengthen the design. In the sections that follow, we present some relevant background on the turnaround intervention, our methods, the findings concerning the impact of turnaround on student outcomes, and our conclusions.

Background

By most measures, the progress in improving student outcomes experienced in North Carolina during the decade from 1991-2000 began to stall at the turn of the century. Due to perceived inequities in education quality across the state, plaintiffs in five rural school districts filed suit against the state in a case known as *Leandro v. State of North Carolina*. The presiding judge, Howard Manning found that high schools in which less than 55% of the students taking state-mandated end-of-course tests (EOC) were passing these tests were committing academic genocide and ruled that the schools should be individually assessed and a plan for improvement developed. Governor Easley later raised the threshold of the PC for state intervention to 60%.

In addition to judicial pressure to improve chronically failing schools, the federal No Child Left Behind Act (NCLB), a reauthorization of the Elementary and Secondary Education Act, sanctioned schools that did not make *Adequate Yearly Progress* for five consecutive years by requiring them to restructure. To provide some context on the magnitude of this provision, by 2010 it was projected that 5% of the nations schools would be sanctioned and required to restructure (AEI & Mass Insight, 2008). In North Carolina, 87 schools were reported to be in the planning or implementation phase of NCLB restructuring during the 2008-09 academic year (CEP, 2009).

NCLB specified five restructuring options: (1) turn the school over to the state, (2) turn the school over to a private education management organization (EMO), (3) reopen as a charter school, (4) replace some or all of the teachers, staff and administrators, or (5) "any other" school restructuring option. Schools in North Carolina, like the majority of schools nationally forced to restructure, chose the fifth option (Le Floch et al., 2006; Mathis, 2009). Because this alternative is "open ended" there has been great variation in the approaches taken across schools, including reopening as a themed school, adopting "schools within a school" models, creating smaller learning communities such as ninth grade academies, developing homegrown reform models, adopting national comprehensive reform models, or a combination of these approaches.

In response, NC DPI established the District and School Transformation Division to create and implement the NC high school turnaround intervention. The high school turnaround intervention required that each low performing high school develop and implement a school improvement plan based on the state prescribed Framework for Action (See Table 1); participate in a centralized program of professional development for the leadership team; receive onsite leadership and instructional coaching and school-specific professional development; and adopt or develop a reform model. As such, we use the term *turnaround* to refer the intervention services NC DPI employed under NCLB's "any other restructuring option." Similar to other studies, turnaround interventions expect quick and drastic school improvements; whereas, school improvement is expected by steady improvements in student performance over three to five years (Kowal & Hassel, 2005; Aladjem et al., 2010; Herman et al., 2008). In total 66 high schools were placed in state turnaround based on scoring below the PC cut-off of 60% for two consecutive years with 35 schools entering state turnaround in 2006-07 and 31 entering in 2007-08.

Table 1: North Carolina's High School Turnaround Strategy

Elements of High School Turnaround Framework for Action (NC Department of Public Instruction, n.d)	
1	Ninth grade transition
2	Formative assessment
3	Meet needs of students who are struggling academically
4	Address literacy issues and needs
5	Professional development based on student achievement data
6	Review all school processes and procedures to ensure that they are structured to help all students achieve proficiency
7	Involve the school community in addressing the needs of the school
8	Establish a professional learning community
9	Determine whether the school will design or adopt reform model

In large measure, the processes and procedures in North Carolina's turnaround intervention correspond to those recommended in the IES's best practice guide.[1] The guide recommended four strategies to turn around low performing schools:

1. "Improved Leadership" to change leadership practices, communicate that dramatic changes will swiftly occur, and replace ineffective teachers.
2. "Quick Wins" to make rapid changes that lead to quick improvements.
3. "Focus on Instruction" by providing targeted professional development, improving instruction, continuously tracking progress and making adjustments.
4. "Committed Staff" where school leadership must build committed staff by assessing the strengths and weakness of their current and newly recruited staff (Herman et al., 2008).

However, based on the standards of the What Works Clearinghouse, the level of evidence for the recommendations in the practice guide was low, which clearly indicates that more research is needed to determine whether school turnaround interventions can be effective at scale.

To justify our attempt to make a causal connection between North Carolina's turnaround intervention and outcomes such as achievement test scores and graduation rates, we examined changes in outcomes for the turnaround high schools PC descriptively. The changes in the PC for the two cohorts of turnaround schools are depicted in Figures 1a and 1b. The change in PC is calculated by mean-centering the PC for each school in the year prior to the turnaround intervention and subtracting that from the centered value for the school in 2009-10, three or four years after the schools entered turnaround, depending on the cohort as shown in equation 1.

---

[1] For a more detailed examination of the turnaround process in these schools see Thompson et al. (2011) and the companion paper presented by Thompson (2012) at the National Center for Scaling Up National Invitational Conference.

(1) $(PC_{j2009-10} - \overline{PC}_{2009-10}) - (PC_{jt-1} - \overline{PC}_{t-1})$,

where $PC_{j2009-10}$ is the achievement test pass rate for school *j* in 2009-10 and PC_{jt-1} is the achievement test pass rate in the year prior to the turnaround intervention; $\overline{PC}_{2009-10}$ is the mean pass rate across all schools in 2009-10 and $\overline{PC}_{t-1}$ is the mean pass rate across all schools in the year prior to turnaround intervention. Centering the data was necessary because the tests included in the PC changed during the study period. The changes in the high school PC was highly variable as shown in Figure 1 with 24 (67%) high schools in the first cohort and 28 (90%) in the second improving their PC while 11 (30%) high schools in the first cohort and 3 (10%) in the second actually lost ground. In the first cohort, 22 exceeded the gains of a comparison group of 64 high schools that were just above eligibility for turnaround in four years and 7 in the second cohort in three years. This suggests that the turnaround intervention may have been effective in some but not in all of the high schools.

Figure 1a: Changes in Performance Composite for Cohort 1 of North Carolina School Turnaround Intervention

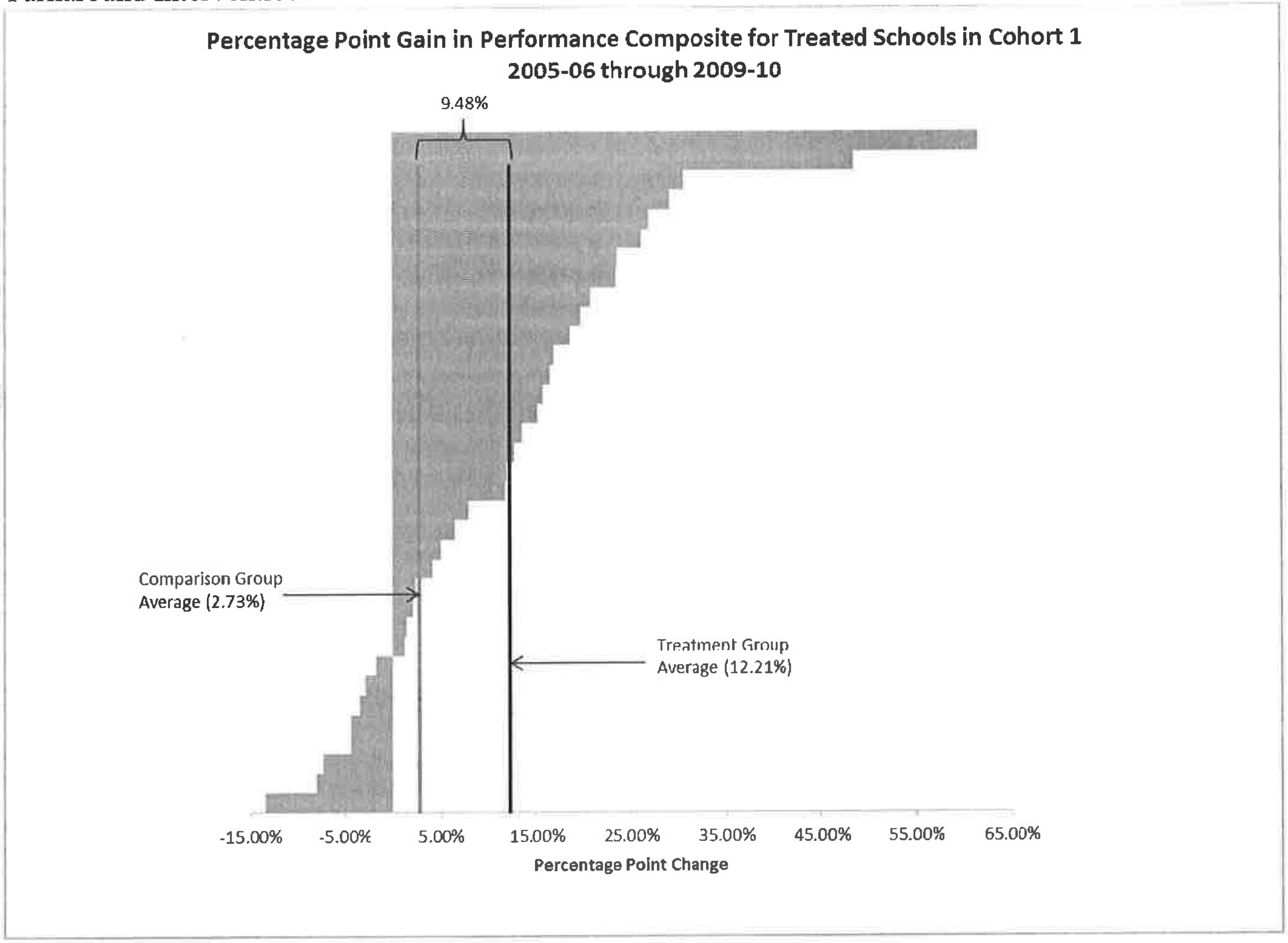
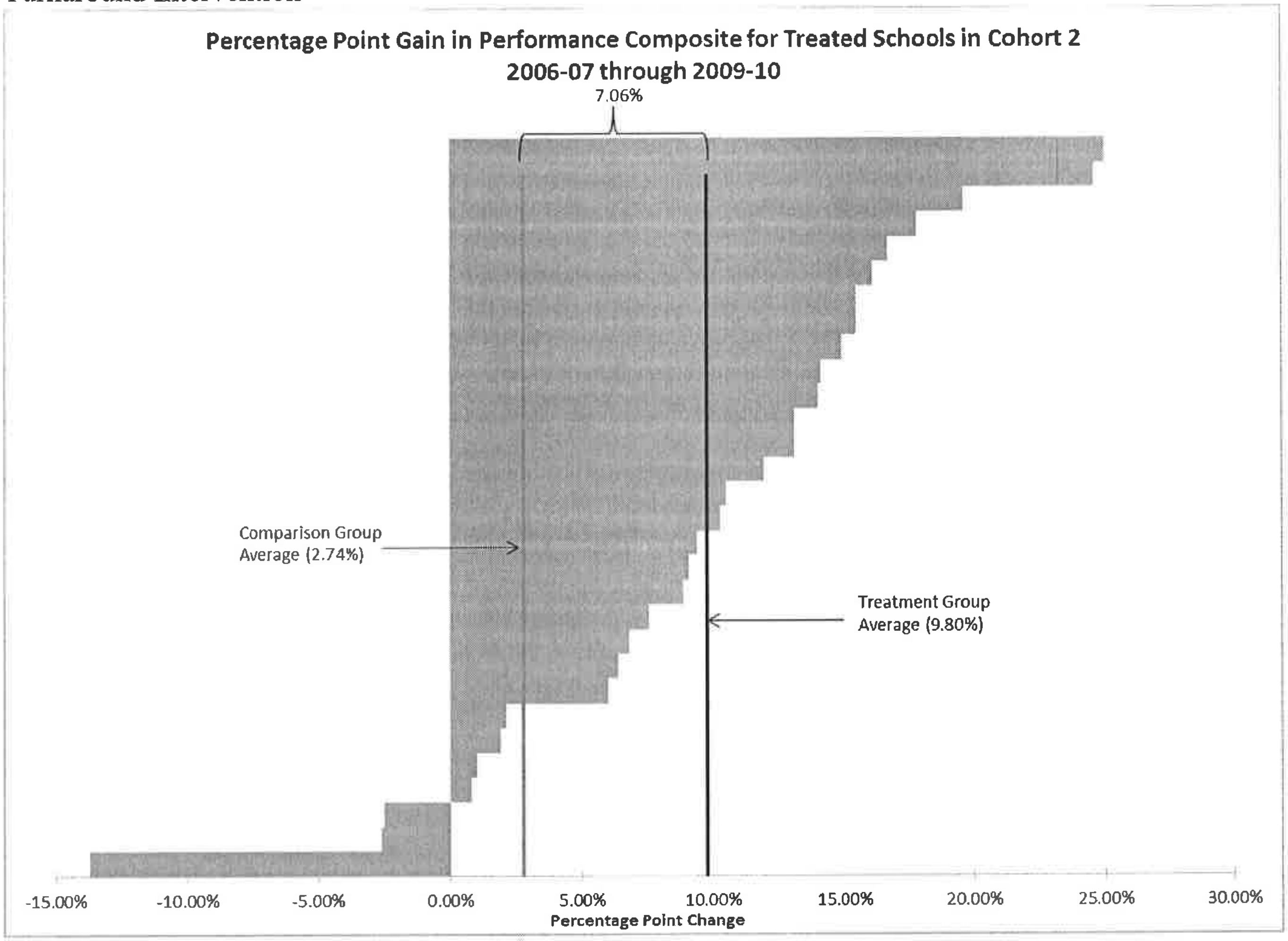


Figure 1b: Changes in Performance Composite for Cohort 2 of North Carolina School Turnaround Intervention



In the next section, we explain the methods that we used to begin to examine the extent to which the change in student outcomes can be attributed to the North Carolina turnaround intervention.

Methods for Identifying and Estimating Turnaround Impact

The goal of the research was to produce an unbiased estimate of the average effect of North Carolina's turnaround approach (labeled "intervention" hereafter) on students' educational outcomes. To realize this goal, random assignment of eligible schools to the intervention (treatment group) or a control group could have produced the most compelling evidence. This option was not available given this study began after the assignment was determined. However, another rigorous identification strategy, regression discontinuity (RD) could be implemented because a *forcing* variable, in this case the PC, was used to assign high schools to the intervention or not. Schools with PCs less than or equal to 60% for two years were intended to be assigned to treatment and those above 60% in either year were not and thus were eligible to be drawn as comparison schools. However, some "crossover" schools that should have been assigned to treatment group did not received intervention and some schools above the cutoff received the intervention. In total the function of PC that minimized crossovers identified an effective cutoff point of 53% in the year prior to being assigned. Using this cutoff point produced 18 crossovers: seven schools below the cutoff failed to receive treatment and eleven schools above the cutoff received treatment. No other single-year or multi-year cutoff point yielded fewer crossovers. Due to the crossovers, we implemented a fuzzy RD (FRD) rather than a sharp RD.

Fully exploiting the existing longitudinal data set, we also chose to use a difference-in-differences (DID) estimation strategy. The DID reduces the possibility that pre-existing differences in the intervention and comparison groups were responsible for the differences in outcomes after the intervention and rules out all events or trends that affected both treated and comparison schools as confounds.

Finally, we used a random effects multilevel model with student-level outcomes at the first level and school outcomes and the indicator for the intervention at the second level since the turnaround intervention was assigned to entire schools. The two-level design maximized power, which is known to be lower for RD designs (Schochet, 2008), and calculated standard errors appropriately for proper inference. Due to the FRD, the forcing variable, PC, was used as an instrument to predict assignment to treatment and two-stage least squares was used for estimation. Several robustness checks along with sensitivity analyses were performed and are discussed further in this paper.

Data. Statewide student test score data collected over a period of six years from 2004-05 to 2009-10 were restructured into pre- and post-treatment periods for turnaround schools in both cohorts. Class rosters provided the link between students and their teachers and classmates in each of their classes for each school year (with a 93% match rate over the six years). Each year of student records was viewed as a serial cross-section for the student-level analysis. Because two cohorts of high schools were admitted to treatment different baseline observations were used. For cohort 1 and non-turnaround schools the school years 2004-05 and 2005-06 comprised the pre-treatment period, and 2006-07 to 2009-10 four post-treatment periods. For cohort 2, 2006-07 comprised an additional pre-treatment period, with 2007-08 to 2009-10 as three post-treatment periods.

Measures. To evaluate the effectiveness of high school turnaround interventions, it is important to include achievement, attainment, and, perhaps, more proximal outcomes. For this study, dependent variables individual EOC test scores in English I, Algebra I, Biology, Civics and U.S. History; all of which were required for high school graduation in North Carolina during the study period. In addition, graduation rate and PC, school-level measures, were examined as student performance. Finally, we included student absenteeism as an outcome based on the expectation that one indicator of positive change could be increased student attendance. Increased attendance could be expected to increase instructional time for students and perhaps signify more positive relationships with students and higher levels of student motivation. Each of the student-level outcome measures were standardized with a mean of zero and standard deviation of one.

The PC, the percent of students at or above grade level on EOC exams, was the "forcing" variable, which assigned schools to the intervention and served as the identifying instrument for assignment to treatment. A rich set of covariates was also included at the student and school-level. Student covariates included 8th grade math and reading end-of-grade test scores; days absent (end-of-course models only); student migration; age-grade discordance; exceptionality; free/reduced price lunch status; race/ethnicity; gender; and current or previous recipient of limited English proficiency services. School-level covariates included proportions of free/reduced price lunch and race/ethnicity; rate of violent acts per 1000 students; average daily membership and its square.

Sample. Of North Carolina's 460 high schools, 130 were included in this study, based on having a 2004-05 or 2005-06 performance composite at or below 66%. This cutoff for the sample was chosen because most of these schools were potentially eligible for school turnaround interventions and taking all high schools would include many with potential ceiling effects on test scores and graduation rates. Among these 130 schools there were a total of 568,043 unique student-by-year records represented. Many student-by-year records were for the same students over time but were treated as serial cross-section (i.e. independent). While failure of independence manifests primarily on reduced standard errors, three factors attenuate our concerns: (1) the overwhelming size of the sample implies that standard errors are likely to be very small in any case; (2) the variables of interest are school-by-student interactions, which are properly inflated through estimation of random intercept models to reflect the design effect within each period; (3) the scores for each of the five EOC tests were estimated in separate models, which minimizes the number of students with multiple observations in any model. For the model of school graduation, there were 622 valid school-by-year observations; the graduation rate variable was not calculated for 2004-05 and therefore, not available.

Identification of Average Treatment Effect. In a FRD, the ratio of the discontinuity in the outcome to the discontinuity in the probability of receiving DST services is the average treatment effect. Using a potential outcomes framework, where we assume that only one of each school's two potential outcomes can be observed, we estimate the treatment impact as the average observed difference between schools assigned to each condition (Morgan & Winship, 2007). Assignment can be represented as a response function conditioned on the cutoff (c) on the forcing variable as follows (with $D_j = \{1$ = assigned to the turnaround intervention; 0 = not assigned to the turnaround inervention$\}$ for school j; X = forcing variable):

$$(2) \qquad E\big(D_j(1)\big|\, X = c\big) - E\big(D_j(0)\big|\, X = c\big)$$

The response of the outcome given X is represented as follows (with Y_j = outcome for school j):

$$(3) \qquad E\big(Y_j(1)\big|\, X = c\big) - E\big(Y_j(0)\big|\, X = c\big)$$

Taking the limits of these expectations in a region ϵ around the cutoff $X = c$, and the ratio of the differences of these limits (provided the difference between the limits in (1) is not equal to zero), gives us the average effect of the turnaround intervention:

$$(3) \qquad \text{ATE} = \frac{\lim_{\epsilon \to 0} E\big(Y_j(1)\big|\, X{=}c{-}\,\epsilon\big) - \lim_{\epsilon \to 0} E\big(Y_j(0)\big|\, X{=}c{+}\,\epsilon\big)}{\lim_{\epsilon \to 0} E\big(D_j(1)\big|\, X{=}c{-}\,\epsilon\big) - \lim_{\epsilon \to 0} E\big(D_j(0)\big|\, X{=}c{+}\,\epsilon\big)}$$

The assumptions needed to identify (3) as the average treatment effect are consistent with the assumptions of local average treatment effects (LATE) (Hahn, Todd & van der Klaaw, 2001;

Lee & Lemieux, 2010). First, the cutoff point can only increase or decrease the probability of assignment to treatment (monotonicity assumption). It cannot do both, and in the present context we assume that it decreases the probability of assignment as we move left to right across the threshold of 53. Second, the forcing variable cannot have an independent effect on the outcome outside of its effect on assignment (exclusion restriction). Intuitively, no school can choose its value on the forcing variable. In other words, each school's value on the forcing variable near the cutoff must be effectively independent. Third, the forcing variable and assignment must demonstrate a positive relationship, the only testable assumption.

Consistent with this interpretation, FRD was implemented using instrumental variable estimation. As recommended by Lee and Lemieux (2010), two stage least squares (2SLS) with a conditional linear probability model for assignment was estimated in the first stage; followed by a second stage modeling the dependent variables on the covariates and the predicted value of assignment.

Difference-in-Difference Estimation of Treatment Impact. To add further consistency to the treatment impact in the context of a panel data set with multiple years' observations, the models incorporate a difference-in-difference estimate by interacting the treatment assignment variable with a *treatment period* variable identifying whether the measure was taken before (pre = 0) or during/after (post = 1) treatment. The difference-in-difference eliminates from the impact estimate any contemporaneous factors that are common to both treatment groups. The interaction is decomposed as follows: first, the treatment assignment indicator captures assignment to the turnaround intervention; the test of the marginal effect for this variable is the absolute difference between the treatment and comparison schools prior to administering the treatment. In contrast to usual comparison group designs, a significant non-zero effect is possible and trivial due to the assignment based on the forcing variable. Second, the marginal effect of the treatment period variable adjusts for changes in the outcome that are common to both comparison and treatment schools; its test (of whether these common changes are significantly different from zero) is trivial. The coefficient for the interaction of these two variables, which represents the incremental impact in the treatment period experienced by schools in the turnaround intervention, is the focal parameter, and its t test is the test of the significance of the LATE of the turnaround intervention.

The difference-in-difference interaction necessitates a minor modification to the instrumental variable framework necessitated by the FRD; a second first-stage regression must be estimated for the product of the period and treatment assignment, with the product of the period and forcing variable as its instrument. The predicted values of the assignment to intervention and period by assignment interaction are then subsequently entered into the second stage regression.

Nested Students-within-Schools Design. The data comprise serial cross-sections of students nested within schools over the period of years examined. The nesting of students within schools implies that a design effect is present for school-level variables, including the treatment indicator, and that standard errors for these parameter estimates should be adjusted accordingly. Therefore, a random intercept model using the XTIVREG procedure in Stata 12 was employed to accommodate random effects instrumental variable designs.

Model. Given the FRD instrumental variable design, difference-in-difference impact estimation with nested data, the models for the treatment and treatment by period first stage regressions for the student-level achievement test scores and absences are:

(1) $$DST_{ij} = \alpha_{00} + X_{ij}\alpha_{xi} + X_j\alpha_{xj} + \alpha_{PC}PC_j + r_{ij} + v_j$$

(2) $$T_tDST_{ij} = \alpha_{t00} + X_{ij}\alpha_{txi} + X_j\alpha_{txj} + \alpha_{tPC}T_tPC_j + r_{tij} + v_{tj}$$

The outcome model (second stage) has the following form, with the instrumented treatment and difference-in-difference variables $\widehat{ST}$and $\widehat{T_1ST}$entered into a model:

(3) $$Y_{tij} = \pi_{00} + X_{ij}\pi_{xi} + X_j\pi_{xj} + \pi_{T1}T_1 + \pi_{ST}\widehat{ST} + \pi_{DID}\widehat{T_1ST} + e_{ij} + u_j$$

The variables X_{ij} and X_j are vectors of student and school characteristics; $T_1$is the post-intervention period dummy variables (reference condition is pre-intervention period); $ST_{ij}$is the school turnaround assignment variable (predicted value of this variable from the first stage replaces the observed data in the second stage); $T_1ST_{ij}$is the interaction between the treatment period dummy variable and the ST assignment variable defining the difference in difference estimator (predicted value of this interaction replaces the observed data in the second stage); PC_j is the forcing variable (instrument), which is a performance composite in the year of assignment; T_tST_{ij} is the interaction between the period dummy variable and performance composite used as the instrument for T_tST_{ij}; r_{ij}, v_j, r_{tij}, v_{tj}, e_{ij} and u_j are errors.

Models for school-level PC and graduation rates was also estimated; further, each of the student-level variables was aggregated to the school level within each year and the model retested with this alternate structure. These models were also multilevel with the "wave" or year for each graduation rate, nested within the school. Aside from the exclusion of covariates for student characteristics and the nesting of time rather than students within schools (that is, the subscript i nested time rather than student), these models were identical to (2) - (4) in form.

Bandwidth. The intuition for RD studies is that at the cutoff the treatment and comparison schools are similar in most respects, and therefore, the difference in the outcomes for the two groups can be reasonably attributed to the treatment. An issue is the choice of the range around the cutoff within which to estimate the LATE. Bandwidth, or the range of the forcing variable around the cutoff used to estimate the LATE, was examined empirically within the treatment and sample of the other low performing schools that we are using for comparisons. Narrower bandwidths focus the analysis near the cutoff point, which is consistent with the intuition. This is desirable in the RD framework, which assumes that the average treatment effect is weighted according to the proximity to the cutoff (Lee & Lemieux, 2010). Farther away from the cutoff point, the potential for bias gets larger, particularly if the functional form of the relationship between the forcing variable and the outcome variable deviates from the functional form that is used in the estimation. However, the preference for narrower bandwidths is counterbalanced on efficiency and by extension power, due to smaller samples of schools available within the specified range. Widening the bandwidth includes more schools but these schools may be dissimilar to schools nearest the cutoff point. We used kernel density estimation and cross-validation techniques to identify likely candidates for optimal bandwidth. The cross validation method examines each dependent variable separately under the assumption that each has its own optimal bandwidth and minimizes the mean squared error (MSE) of prediction based on different bandwidth selections (Imbens & Lemieux, 2007; Lee & Lemieux, 2010). For each outcome variable, we highlight the results that accord with the optimal bandwidth identified by this procedure, but we estimated additional results across a range of half-bandwidths (the partial bandwidth on each side of the cutoff) from 5 to 15 percentage points on the performance composite, as well as results for the full observed range from 15 to 66, as a type of sensitivity analysis.

Additional diagnostics and sensitivity analyses. In addition to the sensitivity analysis for bandwidth described in the previous section, an additional analysis was performed to examine the sensitivity of the results to the specification of the model. Versions of the models without covariates, which are generally regarded as superfluous for unbiasedness in sharp RD studies but may contribute to precision and also lend support to the exclusion restriction in FRD studies, were estimated (Lee & Lemieux, 2010).

Results

Bandwidth Analysis

Kernel density graphs of the treatment groups' overlap on the forcing variable suggest that as the bandwidth is widened it both improves the amount of overlap and reduces the relative density of the comparison group. The reduction in relative density begins after the half bandwidth (h) reaches 13, which is the ceiling on the bandwidth above the cutoff point. A Stata function (kdensity) indicated that the optimal bandwidth for kernel density is 10.

Figure 2: Kernel density for h = 5

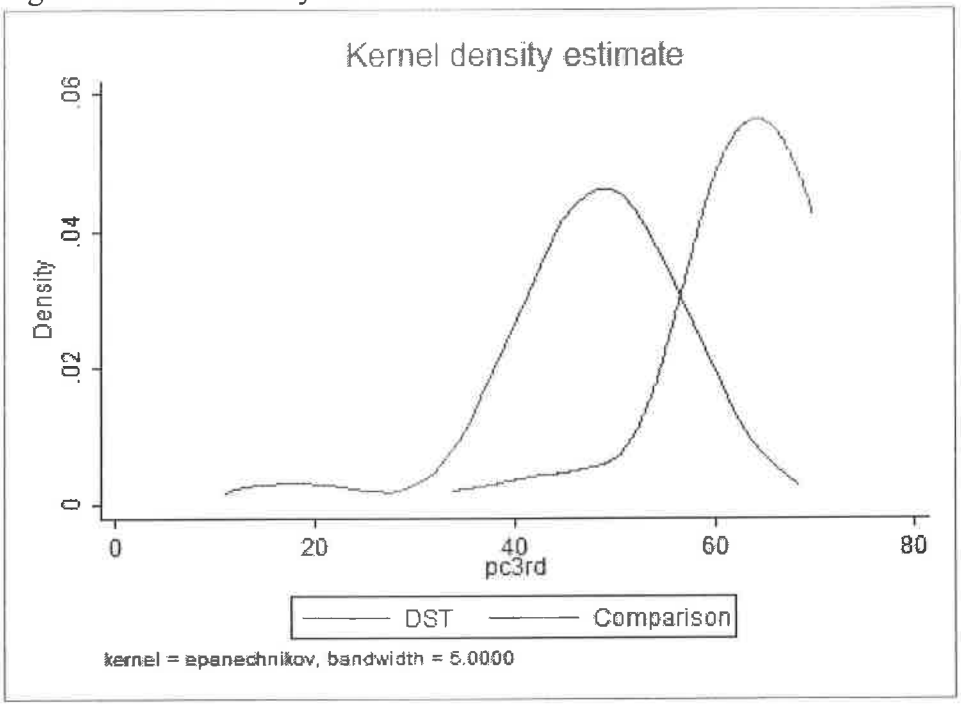


Figure 3: Kernel density for h = 10

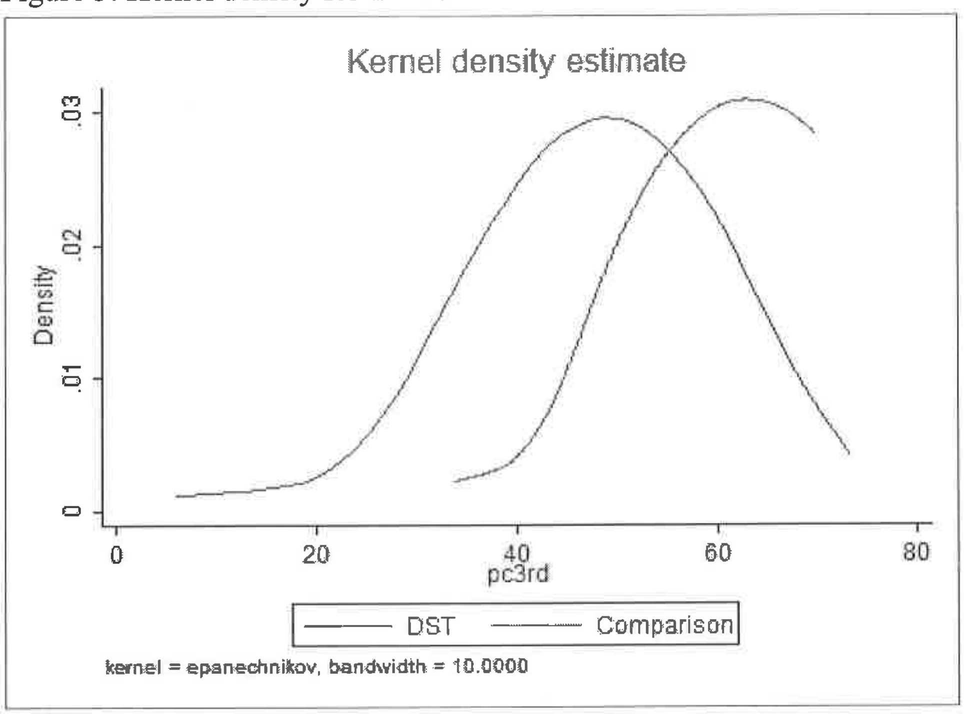
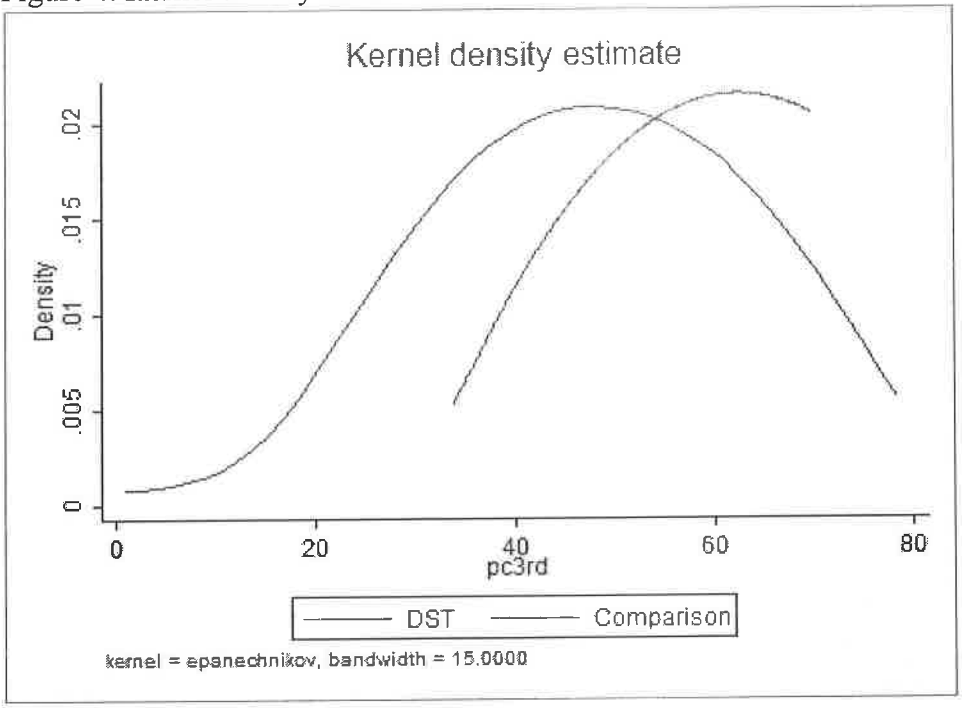


Figure 4: Kernel density for h = 15



Cross-validation techniques that defined an "optimal" bandwidth as one that minimizes MSE demonstrate that the optimal bandwidth varies according to the dependent variable and therefore were estimated separately for each of the five test scores, absenteeism and graduation rates. For Algebra I, the optimal half bandwidth (h*) of those tested was 15; for English I, h* = 9; for Biology, h* = 13; for Civics, h* = 5; for U.S. History, h* = 15; for days absent, h* = 11; and for graduation rate, h* = 11. Because several of the models were found to be optimal near h = 10 (English I; days absent; and graduation rate), which was identified as the optimal bandwidth for kernel density, the MSE for h = 10 was also examined for these models and not found to improve MSE.

Difference-in-Differences Estimates

In Table 2, we report the findings for each dependent variable separately, reporting both the difference-in-difference effect for the model for the full sample (π_{DID}^{F}; 66 turnaround schools and 64 comparison schools) and for the optimal bandwidth for the specific outcome variable (π^{*}_{DID}), and for each of the student-level outcome variables aggregated to the school-level (π^{S}_{DID}). Estimates for the sensitivity test results including variable bandwidth (π^{h}_{DID}, for h from 3 to 15) and the unconditional models (π^{U}_{DID}) are also reported.

Table 2: Estimates of Turnaround Impacts on Student Outcomes

Student Outcomes	Student Level Estimates with Full Sample	Student Level Estimates with Optimal Bin Width	Sensitivity to Bin Widths (Range of Student Level Estimates)	School Level Estimates with Full Sample
Algebra I	.046***	.037***	-.023 – .046	.073
Biology	.066***	.030***	.019 – .066	.063
English I	.019**	.026**	.013 – .043	.067**
Civics	.049***	.113***	.020 – .113	.007
U.S. History	.075***	.074***	.059 – .114	.085*
Student Absenteeism	-.011	-.01	.00 – -.151	-.01
Graduation Rate	NA	NA	NA	.016
Performance Composite	NA	NA	NA	.049***

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Algebra I. As indicated in Table 2, Algebra I test scores were significantly higher in schools that received the turnaround intervention than in the comparisons for the full sample and h* = 15 with coefficients of π^{F}_{DID}= .046, p < .001; π^{*}_{DID}= .037, p < .001, respectively. The effect estimates fluctuate considerably between h = 5 to h = 11, with a sizable and significant impact at the most narrow bandwidths (approximately .04 -.05), becoming negative (-.023) and significant (p < .05) at h = 9. However, the findings for h = 13 and for the full range are similar to that for h*. For the school-level analysis the coefficient was larger but not significant.

English I. For the full sample and h* = 9, test scores for English I were significantly higher in turnaround schools than in the comparison schools ($\pi^F_{DID} = .019$, $p < .001$; $\pi^*_{DID} = .026$, $p < .01$). The effect varies between .013 (h = 11) and .043 (h = 7), but its impact is significant at .01 or better at all values of h and it is significant in the school-level model ($\pi^S_{DID} = .067$, $p < .01$).

Biology. For the full sample and h* = 13, test scores for Biology were significantly higher in turnaround schools than in the schools without turnaround ($\pi^F_{DID} = .066$, $p < .001$; $\pi^*_{DID} = .030$, $p < .001$). However, the difference in Biology test scores between the turnaround schools and the comparison schools persists across the values of h tested, though the effect is lower at h = 9, and inflated at h = 5 and 7. It is not significant at the .05 level at h = 9. At the school-level the effect is about the same magnitude but no longer significant ($p < .05$).

Civics. For the full sample and h* = 5, Civics test scores were significantly higher in the turnaround schools than in comparison schools ($\pi^F_{DID} = .049$, $p < .001$; $\pi^*_{DID} = .113$, $p < .001$). The effect is much lower at all other values of h (in a range of .034 - .046), but significant in each comparison except at h = 7. At the school-level, the turnaround school effect estimate is insignificant.

U.S. History. For the full sample and the optimal bandwidth, h* = 15, U.S. History scores were significantly higher in turnaround schools than comparison schools ($\pi^F_{DID} = .075$, $p < .001$; $\pi^*_{DID} = .074$, $p < .001$). U.S. History shows a slightly inflated effect at lower bandwidths, but has a stable estimate of .07-.08 at half bandwidths of 11 or higher. It is always significant at .01 or better. The magnitude of the school-level effect is similar and statistically significant ($p < .05$).

Student Absences. For the full sample and h* = 11, student absenteeism is lower in turnaround schools than in comparison schools, but not significantly. Only at h = 5 and h = 7 does absenteeism demonstrate a significant effect (π_{DID}^{5} = -.151, $p < .001$; π_{DID}^{7} = -.032, $p < .05$). The effect is attenuated at higher bandwidths and in the school-level model and are not significant.

Graduation Rate. For the full sample and for all h, including the optimal bandwidth, h* = 11, school graduation rate is positive but not significantly affected by the turnaround intervention. While the effect sizes are somewhat smaller than those observed for many of the student-level models, the smaller sample size of the school-level models yield higher standard errors.

Performance Composite. For the full sample, the PC was positively and significantly affected by the turnaround intervention. The effect size was .05 for the full sample, indicating that the broadest measure of school performance was moved in the direction consistent with positive effects of the turnaround intervention.

Sensitivity Analyses

The optimal bandwidth models for each dependent variable were re-estimated using unconditional models without covariates (only the period, treatment and period by treatment difference-in-difference were estimated, specified as π_{DID}^{U}). For Algebra I, the effect was largely unaltered (π_{DID}^{U} = .034; $p < .001$); for English I, the effect was much larger (π_{DID}^{U} = .089; $p < .001$); for Biology as well the effect was much larger (π_{DID}^{U} = .071; $p < .001$); for Civics, the effect was slightly lower (π_{DID}^{U} = .095; $p < .05$); for U.S. History the effect was slightly larger (π_{DID}^{U} = .087; $p < .001$); for student absenteeism the effect was smaller and remains insignificant as in the models conditioned on covariates (π^{U}_{DID}=.0113); for graduation the effect was similarly insignificant (π^{U}_{DID} = .010).

In summary, the turnaround intervention appeared to improved test scores in the five core courses required for graduation—Algebra I, English I, Biology, Civics and U.S. History as well as the overall Performance Composite– but the effects on student absenteeism and graduation rates, while in the direction which would indicate positive effects, were not significant.

Conclusion

The North Carolina turnaround school intervention was effective in raising test scores and schools overall PC but not raising graduation rates or lowering student absenteeism. The descriptive analysis shows that the effects are heterogeneous, which is also consistent with the fluctuations in the sizes of the effects and significance across the bandwidths assessed: as the bandwidth gets wider, the sample of schools changes and the heterogeneity of effects between schools suggests we should see changes in the size and significance of the impact.

The effect estimates strongly suggest that the North Carolina turnaround schools, intervention was a success but the heterogeneity also suggests a plausible hypothesis for further investigation – that more high schools may be turned around by improving implementation. It is unclear from this analysis whether the fidelity of implementation needs to be improved or if the effects are moderated in ways that suggest that the intervention should be modified in certain types of schools or with particular groups of students. The qualitative research on the North Carolina turnaround intervention suggests that district-level support for the turnaround reforms are essential and this may be a moderating factor. For example, we can imagine districts that support the turnaround might place, or provide incentives that encourage, the best available teachers and principals to move to the turnaround schools and support decisions that principals make that may be unpopular with parents or teachers but are likely to improve instruction and student outcomes. Because district support is not currently systematically measured, the district support hypothesis cannot be tested directly with the available data.

While the turnaround model has been shown to be successful, it remains unclear what this means for efforts to implement the type of school transformations called for in more recent policy directives such as Race to the Top (RttT). As a part of the American Recovery and Reinvestment Act (ARRA), RttT provides monetary incentives to those states that receive RttT funding to undertake initiatives designed to improve performance in the lowest achieving schools and districts. Like the NCLB, RttT specifies models to be followed for improving these school but the RttT models are different than those prescribed in NCLB. The four RttT interventions are: (1) turnaround, (2) restart, (3) transformation, or (4) school closure (see Table 1 for a detailed description of the restructuring models). The RttT prescribed model for school reform that is closest to the NC turnaround intervention is turnaround but there are important differences, such as the replacement of administrators and staff.

References

Aladjem, D.K., Birman, B.F., Orland, M., Harr-Robins, J., Heredia, A., Parrish, T.B., & Ruffini, S.J. (2010). U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, Achieving Dramatic School Improvement: An Exploratory Study, Washington, D.C.

Balfanz, R. & Legters, N. (2004). Locating the dropout crisis. Which High Schools Produce the Nation's Dropouts? Where Are They Located? Who Attends Them? Baltimore, MD: Center for Research on the Education of Students Placed At Risk. Report No. 40.

Balfanz, R., Bridgeland, J., Moore, L., & Fox, J. (2010). Building a grad nation: Progress and challenge in ending the high school dropout epidemic. Civic Enterprises Everyone Graduates Center at Johns Hopkins University. Baltimore , MD.

Borman, G.D., Hewes, G. M., Overman, L.T. & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research* 73(2), 125-230.

Center on Education Policy (2009). Improving low-performing schools. Lessons from five years of studying school restructuring under *No Child Left Behind.* Washington, D.C.

Imbens, G. W. & Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615-635.

Hahn, J. Todd, P. & van der Klaaw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., & Darwin, M. (2008). Turning around chronically low-performing schools: A practice guide (NCEE #2008-4020). Washington, DC: 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/publications/practiceguides.

Kowal, J.M. & Hassel, E.A. (2005). School restructuring options under No Child Left Behind: What works when? Turnarounds with new leaders and new staff. The Center for Comprehensive School Reform and Improvement. Retrieved from http://www.centerforcsri.org

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

Le Floch, K. C., Taylor, J., & Zhang, Y. (2006). Schools in NCLB restructuring: National trends. Washington, DC: American Institutes for Research.

Lemann, N. (2010, September 27). Schoolwork. *The New Yorker.*

American Enterprise Institute & Mass Insight Education and Research Institute (2008). "Turning Around the Nation's Worst Schools. Papers prepared for the American Enterprise Institute/Mass Insight Education & Research Institute Conference

Mathis, W. (2009). NCLB's Ultimate Restructuring Alternatives: Do they Improve the Quality of Education? Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit

Morgan, S. L. & Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. NY: Cambridge University Press.

Murphy, J. (2009). Turning around failing schools: Policy insights from the corporate, government, and nonprofit sector. *Educational Policy*, 23(6), 796-830.

Schochet, P. (2008). Technical Methods Report: Statistical Power for Regression Discontinuity Designs in Education Evaluations. Washington, DC: Institute of Education Sciences.