Low-Performing Schools and School Reform: Three Essays on School Turnaround, the Mechanisms of Low Performance, and Leadership for Reform

By

Erica Harbatkin

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

June 30, 2020

Nashville, Tennessee

Approved:

Gary T. Henry, Ph.D.

Shaun M. Dougherty, Ed.D.

Jason A. Grissom, Ph.D.

John P. Papay, Ed.D.

Kristopher J. Preacher, Ph.D.

For Sydney, who didn't make this process any easier.

And for David, who did.

Acknowledgements

I am grateful to my committee, and in particular to my chair, Gary Henry, for providing guidance and multiple rounds of feedback on this dissertation—especially as they all surely had more pressing things to do during a pandemic.

Thanks to my cohort—in particular to Clay Francis, for helping to get me through the first two years of this process, and to Josh Bleiberg, for helping me through the rest of it.

Thanks to Nicole Arshan, without whom I probably wouldn't have embarked on this Ph.D., and who has been a sounding board and cheerleader over the last five years.

Thanks to Michael Wiseman, who provided mentorship during my master's degree and continued to support me as I moved forward in my career. He didn't quite get to see me cross the finish line but I'm immensely grateful for the support he provided up to the end.

And to David, who took on the bulk of the childcare so I could focus on my work while dudes covered in the New York Times and Washington Post locked themselves in home offices and grumbled about the noise as their wives sacrificed their careers and their sanity. Who followed me from D.C. to Nashville and now to Lansing, MI, and barely complained at all. I can't imagine having done any of this without you, and I'm glad I didn't have to.

Page

Acknowledgements iv

List of Tables vii

List of Figures x

Introduction 1

Chapter

1 The Next Generation of State Reforms to Improve their Lowest Performing Schools: An Evaluation of North Carolina's School Transformation Initiative 2

School turnaround 4
North Carolina Transformation.. 6
Sample.. 11
Data and Measures 13
Outcome measures 13
Assignment variable.. 13
Teacher effectiveness 14
Implementation 15
Covariates 17
Empirical Strategy 18
Main effects 18
Teacher effectiveness 21
Implementation 24
Results.. 26
Validity checks... 41
Discussion 44
Conclusion 47
References 49
Appendix.. 52

2 An Early Warning System for Low-Performing Schools: Developing a Multidimensional Measure of Risk 67

Literature Review.. 68
School turnaround and improvement... 68
School performance and decline 74
Theoretical framework.. 79
Methods.. 81
Data and sample 81
Measures 82
Empirical strategy 87
Results.. 93

School performance 93
Exploratory analysis 108
Measurement model 113
Prediction 116
Discussion 118
References 122
Appendix 127

3 Are There Turnaround Principals? 129

Literature Review 131
School turnaround 131
Principal effectiveness 133
Turnaround principals 135
Methods 138
Data and sample 138
Measures 139
Empirical strategy 155
Results 161
RQ1: How frequently are principals successful in low-performing schools and how often do they sustain that success over multiple years? 161
RQ2: Do successful principals of low-performing schools leave at higher rates than successful principals of non-low-performing schools? 161
RQ3: Where do principals with demonstrated prior success in low-performing schools ("turnaround principals") go after leaving the formerly low performing school? 166
RQ4: Are these turnaround principals more effective at leading low-performing schools than principals who have successfully led non-low-performing schools? 170
Robustness checks 177
Discussion 177
References 180
Appendix 186

List of Tables

Table Page

1-1. CNA and unpacking timing ... 8

1-2. Coaching visits ... 10

1-3. School sample characteristics ... 12

1-4. ITT estimates *(outcome=test score)* ... 28

1-5. TOT estimates *(outcome=test scores)* ... 31

1-6. TOT estimates *(outcome=teacher turnover)* ... 34

1-7. TOT estimates on teacher turnover and new-to-school teachers by lagged teacher effectiveness ... 36

1-8. Sample balance on standardized variables, conditional on forcing variable within optimal bandwidths ... 43

1-9. Attrition ... 44

A-1-1. Examination of treatment receipt by compliance ... 52

A-1-2. Tests for validity of implementation groupings ... 53

A-1-3. TOT estimates within alternative bandwidths and full sample *(outcome=test score)* ... 55

A-1-4. TOT estimates by subject *(outcome=test score)* ... 56

A-1-5. TOT estimates without lagged test score *(outcome=test score levels)* ... 57

A-1-6. TOT estimates by school level *(outcome=test score)* ... 58

A-1-7. TOT estimates within alternative bandwidths and full sample *(outcome=teacher turnover)* ... 60

A-1-8. ITT estimates *(outcome=teacher turnover)* ... 61

A-1-9. Placebo estimates from fuzzy RD within optimal CCT bandwidth, 2016 *(outcome=test score)* ... 62

A-1-10. School demographics by treatment year ... 63

A-1-11. Fuzzy RD results by CNA timing *(outcome=test score)* ... 63

2-1. Components and weights of ESSA-compliant school performance measure 82

2-2. Variables Tested As Predictors of Low Performance 83

2-3. Number of years in the bottom 5%, 2010-11 through 2016-17.. 96

2-4. Prior year percentile of bottom 5 percent, 2011-12 through 2017-18.................................. 97

2-5. Confusion matrix (main index).. 110

2-6. Predictors of low performance 111

2-7. Post-lasso regression results predicting performance level on three performance indices (standardized coefficients).. 112

2-8. Information criteria by number of retained factors 114

2-9. Rotated factor loadings.. 115

2-10. Model fit by school urbanicity.. 116

2-11. Model predictions for year *t*+1, *t*+2, and *t*+3 (main index)... 117

A-2-1. Random partition crossvalidation summary statistics... 127

A-2-2. Rotated factor loadings (orthogonal rotation) 128

3-1. Connected network sizes 143

3-2. Frequency of effective principal by low-performing status in principal's first year in a school, 2010-2018... 145

3-3. Frequency of effective principal within first two years in a school by low-performing status in principal's first year in a school, 2010-2018 146

3-4. Frequency of effective principal for 1-3 years by low-performing status in principal's first year in a school, 2010-2018.. 148

3-5. Sending school frequency of turnaround and effective principals by definition and *sending* school status in principal's first year, 2010-2018... 149

3-6. Receiving school frequency of turnaround principals and other effective principals in a low-performing school by definition, 2011-2018 152

3-7.1 Attrition rates by effectiveness definition and school low-performing status (LP=bottom 10%)... 162

3-7.2 Attrition rates by effectiveness definition and school low-performing status (LP=bottom 25%)... 163

3-8. Probability of attrition by school low-performing status and principal effectiveness by definition 165

3-9. Pathway out by principal effectiveness and low-performing status 167

3-10. Pathway out for movers by turnaround principal category 170

3-11. Turnaround and effective principal effects on student achievement 172

3-12. Turnaround and effective principal effects on teacher turnover 174

3-13. Turnaround and effective principal effects on chronic absenteeism 176

A-3-1. Turnaround and effective principal effects on math scores 186

A-3-2. Turnaround and effective principal effects on reading scores 187

A-3-3. Balance checks on turnaround and other effective principals in receiving schools 188

List of Figures

Figure Page

1-1. North Carolina Transformation logic model 7

1-2. Student achievement by distance from assignment threshold 27

1-3. Proportion treated by forcing variable 29

1-4. Teacher turnover by distance from assignment threshold 33

1-5. Heterogeneity of effects by Comprehensive Needs Assessment timing 39

A-1-1. Graphical integrity of the forcing variable 64

A-1-2. Estimated effects by presence of CNA unpacking 65

A-1-3. Estimated effects by coaching dosage 66

2-1. Path Diagram for measurement model 92

2-2. Bottom 5 percent of schools in North Carolina by year since Race to the Top 95

2-3. Prior year performance of bottom 5 percent of schools 99

2-4. One-year performance change of schools across the full spectrum of school performance 101

2-5. One-year performance change for bottom 5% of schools in year *t* 103

2-6. Trajectories of a random sample of low-performing schools 105

2-7. Mean recall rate for each λ value over 1,000 bootstrapped samples each 109

2-8. Screeplot 114

3-1. Distribution of principal value-added by school low-performing status 141

3-2. Definitions of turnaround principal 147

3-3. Pathway out by principal effectiveness definition B and school low-performing status (LP=bottom 10%) 169

Introduction

While federal education policy requires that states support their lowest performing schools, the research on how to effectively turn around these schools is limited. Drawing from 13 years of statewide administrative data combined with survey data from the North Carolina Teacher Working Conditions survey, this dissertation examines three dimensions of school turnaround and low-performing schools. The first essay is an evaluation of the North Carolina Transformation initiative, an intervention to turn around 75 of the state's lowest performing schools. Using a regression discontinuity design, I find that the intervention decreased student achievement and increased teacher turnover in treatment schools. I find no evidence that the increased teacher turnover was strategic on the part of school leaders. The negative effects on student achievement appear to be related to the timing of the needs assessment, a requirement under federal policy for supporting low-performing schools. The second essay aims to develop an early warning system for low-performing schools by identifying the leading indicators of low performance. Using lasso regression, I find that a small set of indicators related to student preparedness, teachers, and school conditions can accurately predict performance in about 94 percent of all schools and low performance for about 37 percent of low-performing schools. The third essay investigates the phenomenon of so-called turnaround principals. Using a drift-adjusted value-added approach to measuring principal effectiveness, I find that there are principals who are effective in low-performing schools and who sustain that effectiveness over multiple years. However, school fixed effects models do not find evidence that effectiveness transfers to subsequent low-performing schools. Taken together, these three essays provide context for understanding the determinants of low performance and the path forward for states supporting their lowest performing schools under the Every Student Succeeds Act.

Chapter 1

The Next Generation of State Reforms to Improve their Lowest Performing Schools: An Evaluation of North Carolina's School Transformation Initiative

The mandate for continuous support and improvement of each state's lowest performing schools along with the accountability requirements in the Every Student Succeeds Act (ESSA, 2015) will ensure that every state will continue to identify and attempt to reform its lowest performing schools into the foreseeable future. State turnaround interventions under prior federal programs, School Improvement Grants (SIG) and Race to the Top (RttT), have shown some evidence of positive effects on student outcomes (Carlson & Lavertu, 2018; Dee, 2012; Papay & Hannon, 2018; Sun et al., 2017; Zimmer et al., 2017), although some studies have found negative or null effects (Dickey-Griffith, 2013; Dragoset et al., 2017; Heissel & Ladd, 2018; Henry et al., 2015). In many of the turnaround efforts that have been shown to be effective, strategic staffing—which involves replacing less effective teachers by recruiting, hiring, developing and retaining more effective teachers—seems to have played a role in successful turnaround, as we discuss later. However, reforms that successfully recruit and retain effective teachers from other schools may also induce general equilibrium effects, which lowers performance in the schools from which the teachers transferred when the supply of effective teachers is limited (Kho et al., 2019).

Under ESSA, the federally mandated turnaround models have faded into the past along with additional dedicated funding for turnaround. The school reform interventions implemented under No Child Left Behind (NCLB) waivers yielded less consistent effects on student achievement, with just one study finding positive effects, three with null effects, and one with negative effects (Atchison, 2020; Bonilla & Dee, 2017; Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2017, 2018). Within the same intervention, heterogeneous effects have been driven by variation in implementation of reforms (Dougherty & Weiner, 2017; Strunk et al., 2016).

Turnaround under ESSA will share more in common with NCLB waivers and similar district- and state-initiated reforms than RttT and SIG for two reasons. First, states will have flexibility in how they reform their lowest performing schools rather than being required to follow a federally prescribed model. Second, states will undertake turnaround without the infusion of additional federal funds that characterized RttT and SIG reforms. One state-initiated reform operating in this context was the North Carolina Transformation (NCT) initiative, which began in 2015 after the state's services under RttT ended. This study examines the effects of this new round of school support on student achievement and teacher turnover. We ask four research questions:

1. What is the effect of the efforts to improve the lowest performing schools on student achievement?
2. What is the effect of the efforts to improve the lowest performing schools on teacher turnover?
3. Did the reform schools engage in strategic replacement of teachers by hiring more effective replacement teachers and losing less effective teachers?
4. Is variation in implementation associated with differences in outcomes?

By way of preview, relying upon a rigorous regression discontinuity design, we find negative effects on student achievement gains and increased teacher turnover in the second year of services. Moreover, exiting teachers were not less effective than those exiting control schools, and entering teachers were not more effective than those entering control schools, undermining the strategic replacement hypotheses. The negative effects appear to be associated with the timing of the comprehensive needs assessment, a required component of reform under ESSA, which was intended to precede the state's supports to its lowest performing schools. These findings may serve as a cautionary tale for how states engage in mandated comprehensive school improvement (CSI) under ESSA.

School turnaround

Prior research has shown substantial heterogeneity in the effects of whole school reform efforts (Gross et al., 2009). SIG and RttT introduced school turnaround to the federal school reform agenda in 2008. Turnaround was distinguished by the urgency to create dramatic and rapid change in chronically low-performing schools (Herman et al., 2008; Peurach & Neumerski, 2015). Unlike the prior incremental school reforms under CSRD, RttT and SIG required specific practices for disrupting the status quo as part of federally mandated turnaround models. These intentional disruptions included practices such as replacing the principal, replacing at least 50 percent of staff, or restarting the school under new management to allow complete staff replacement (see, e.g., Zimmer et al., 2017). Turnaround efforts funded through RttT and SIG as well as reforms following similar models—many of which included substantial staff replacement and practices aimed at recruiting, retaining and developing effective teachers—produced strong positive effects on student achievement in Massachusetts, Tennessee (local Innovation Zones), Ohio, and California (Carlson & Lavertu, 2018; Dee, 2012; Henry et al., 2020; Papay & Hannon, 2018; Schueler et al., 2016; Strunk et al., 2016; Sun et al., 2017; Zimmer et al., 2017). The mostly positive effects have largely dominated the conversation about turnaround under RttT and SIG, but the average and local average treatment effects mask heterogeneity within interventions. Turnaround in North Carolina, Tennessee, and Texas produced mixed effects (Dickey-Griffith, 2013; Heissel & Ladd, 2018; Zimmer et al., 2017), and some of the interventions yielding positive effects also produced null or negative effects in particular contexts (Carlson & Lavertu, 2018; Strunk et al., 2016; Zimmer et al., 2017).

Heterogeneity of the effects of school reform models continued under NCLB waivers, with fewer positive effects than RttT and SIG. Of four states with evaluations of waiver reforms, one—Kentucky—produced positive effects on student achievement, which the authors attributed to the state's focus on reducing achievement gaps combined with a clearly articulated set of reform activities from the state (Bonilla & Dee, 2017). Reforms in New York, Michigan, Rhode Island, and Louisiana produced either null or negative effects on student achievement (Atchison, 2020; Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2017, 2018).

The mixed effects of interventions under both RttT/SIG, NCLB waivers, and locally initiated reforms underscore three important conclusions about school reform. First, recruiting and retaining effective teachers appears to be a key strategy for achieving and sustaining turnaround. Second, successfully shifting the climate and daily operations of an underperforming school may require some disruption of the status quo. And finally, the impacts of school reform interventions are not universally positive or even neutral—these interventions have the potential to do harm, as they did in some schools in Los Angeles, Rhode Island, North Carolina, New York, Texas, and Michigan.

This paper proceeds as follows. In the next section, we describe the intervention and theory of change under NCT and provide some context on implementation. We then describe the sample, data and measures, and empirical strategy, followed by the findings, and a series of validity checks. We conclude with a discussion of the relevance and limitations of these findings for future school turnaround.

North Carolina Transformation

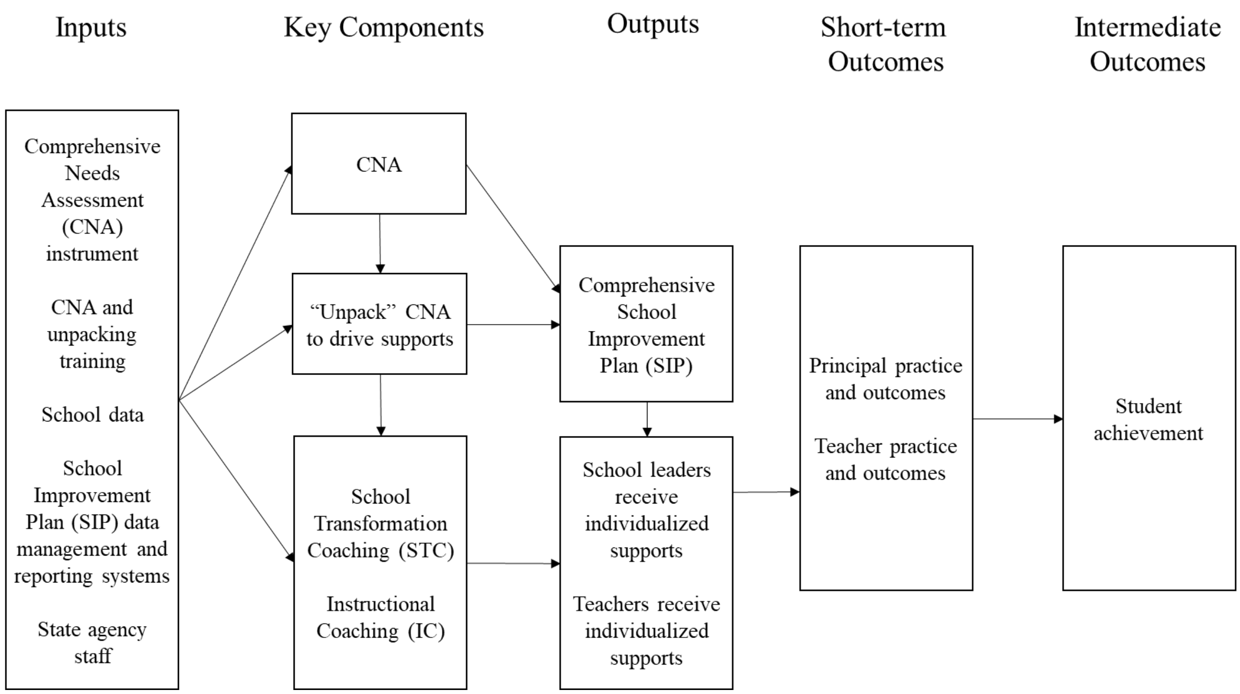
NCT began during the 2015-16 academic year and was implemented in 75 low-performing schools over two academic years. NCT schools received coaching and support services directly from the state Department of Public Instruction (DPI), which had carried out two prior rounds of school turnaround interventions. The most recent was Turning Around the Lowest Achieving Schools (TALAS), the state's RttT turnaround intervention, which focused on reforming 118 schools under the closure (12 schools), transformation (93 schools) and turnaround (14 schools) models through direct service provision from the state Department of Public Instruction (DPI) (Henry et al., 2015). Under TALAS, schools received district-level, school-level, and instructional coaching from about 150 coaches (Henry et al., 2014). All schools in the bottom 5 percent of the state based on the 2009-10 proficiency rate received services. When services ended, a leaner DPI set out to continue its work in a smaller group of low-performing schools.[1] An early adopter of turnaround because of a 2006 court order, North Carolina continued its turnaround efforts without the federal pressures that motivated waiver-based reforms during the same time period.

NCT followed a similar direct services model to TALAS but the selection process excluded schools in the 10 largest districts in the state. As a result, NCT schools were largely rural and, on average, higher performing than TALAS schools. NCT also didn't include require implementation of one of the four federal turnaround models or the federally recommended practices. The NCT theory of action as depicted in Figure 1-1 began with a Comprehensive Needs Assessment (CNA) in which DPI staff would spend two days at treatment schools collecting data through classroom observations, interviews, and focus groups. The state

[1] 21 of the 75 NCT schools had received services under TALAS.

prioritized conducting CNAs in NCT schools that had not received one in the three years prior to the intervention. Of the 75 NCT schools, 84 percent received a CNA prior to this round of school reform or during the two academic years in which services were delivered (See Table 1-1 for timing of CNAs).

Figure 1-1. North Carolina Transformation logic model



Following the CNA, the NCT model called for "unpacking" the findings by state facilitators who discussed CNA findings with school staff. The 1.5-day unpacking process consisted of three elements: (1) the facilitators reviewed the full CNA report with attendees, (2) the facilitators and school staff carried out a "root-cause analysis" in which they sought to uncover the underlying causes of the issues identified in the CNA, and (3) the facilitator and school staff engaged in a planning activity that involved visually mapping the school improvement process moving forward. The unpackings generally occurred during the summer following the school year of the CNA, although there was variation in when and whether schools received them.

Table 1-1. CNA and unpacking timing

	Number of schools	
	CNA	Unpacking
2014-2015	17	15
Spring 2016	25	4
Summer 2016	0	24
Fall 2016	15	1
Spring 2017	4	2
Summer 2017	0	3
Fall 2017	2	0
Pending	0	4
None during intervention period	12[a]	22[b]
Total schools:	*75*	

[a] Of these 12 schools that did not receive CNAs, four declined and eight were not conducted due to Hurricane Matthew.
[b] Of these 22 schools that did not receive unpackings, 12 were schools without CNAs, two declined, two were schools that had received CNAs in fall 2017 because they were under consideration for the state's Innovative School District (ISD), and the remaining six were not conducted for unknown reasons.

The CNA and unpacking were intended as the springboard from which school improvement planning and subsequent turnaround activities would occur. Early in the school year, all low-performing schools in North Carolina were required to submit a School Improvement Plan (SIP) in which priority areas and goals were intended to be based in part on CNA findings. To that end, the timing of the CNA was central to each school's turnaround plans. Because CNA findings were intended to inform the SIP, the NCT model relied on the CNA occurring prior to the school improvement planning process. However, planning delays combined with limited state resources precluded the state from carrying out CNAs in a manner consistent with the theory of action. Only about one-fifth of treatment schools received CNAs before developing their SIPs, though these occurred as many as three years prior to the intervention when schools could have had different staff and different needs. Meanwhile, one-third of treatment schools received CNAs during the spring semester of the first year of the intervention after they had already developed their SIPs.

The CNA, unpacking, and SIP comprised the foundation for turnaround. This framework parallels ESSA requirements, which call for districts to work with low-performing schools to develop a comprehensive support and improvement plan using data from a school-level needs assessment and for the state to monitor school progress on that plan. The core of the improvement intervention was the coaching that followed, with the goal of building leadership capacity through school transformation coaching and teaching capacity through instructional coaching. NCT was intended as a tailored intervention in which coaches were responsive to school, principal, and teacher needs. Not all schools received both school transformation and instructional coaching, and there was wide variation in the number, content, and structure of the visits. While the average treatment school received 45 instructional coach visits and 25 school transformation coach visits over the three-semester period, Table 1-2 highlights the considerable variation across schools. Of particular relevance is the wide range of dosage; treatment schools receiving very few coaching visits experienced fundamentally different exposure to services than treatment schools receiving weekly visits.

Table 1-2. Coaching visits

	Instructional	School transformation
Total schools with coaches assigned	65 total 16 math, 18 ELA, 12 science, 33 non-subject-specific	56 total
Number of visits	Range: 0-137 Mean: 45.36	Range: 0-63 Mean: 25.28
Visits per teacher	Range: 0-15.75 Mean: 1.83	Range: 0-3.82 Mean: 1.03

Source: DPI coaching reports for three semesters from spring 2016, fall 2017, and spring 2017.
NOTE: Subject-level and non-subject-specific ICs do not add up to 65 because schools have ICs focused on multiple subjects. Means are for all treatment schools regardless of whether they have a coach assigned. Visits per teacher based on number of FTE teachers employed in the school across all treatment schools.

The intervention did not closely mirror any of the four previous federal school turnaround models. Instead, the NCT theory of change focused on building staff capacity and gave districts autonomy to transform their low-performing schools using locally developed strategies. In its focus on instructional quality, NCT, like RttT and SIG, recognized the importance of highly effective teachers to school turnaround but focused resources on developing existing staff rather than on recruiting and retaining effective staff. While NCT served the state's low-performing schools during the period between RttT and ESSA, the model aligns more closely with ESSA's flexible approach to school turnaround than with the prescriptive "top-down" turnaround models. This evaluation can therefore help to inform state turnaround policy under ESSA, under which states are required to undertake needs assessments and school improvement plans in their lowest performing schools and have the flexibility to implement school turnaround interventions that look like NCT.[2]

[2] While states may choose to follow school reform models that parallel the four RttT/SIG models, a separate analysis of all state ESSA plans shows very few states have committed to doing so. A total of five states outlined policies in their ESSA plans that committed to state takeover, transferring low-performing schools to alternative management, or staff replacement.

Sample

The sample includes all North Carolina schools that the state determined were eligible for treatment based on data from the 2014-15 school year. Schools were excluded from eligibility for services if they had a school performance grade of C or above for the 2014-15 school year, exceeded growth standards, were situated in one of the 10 largest school districts in the state, or in Halifax County, which was targeted for a district-level turnaround from 2009-10 through 2016-17. Special schools, charter schools, and freshman academies were also excluded. In total, 331 schools were eligible for services and 78 were assigned to treatment. Noncompliance occurred on both sides of the treatment cutoff because state officials did not serve schools without district agreement. In some cases, district officials requested that the state deliver services to a school above the cutoff rather than the school selected, or requested that a particular school be served in addition to the targeted schools. In order to mitigate bias that would arise from these always-takers and never-takers, our inferences apply only to compliers. Sixty-nine of the 78 schools below the cutoff complied with their assignment, nine below the cutoff declined, and six schools above the cutoff received services.

Of the 78 schools below the assignment threshold, 72 were rural, five were in towns, and one was in a city. On average, treatment schools had higher rates of minority and low-income students, higher rates of novice teachers, higher per pupil spending, and lower enrollment than other eligible schools, which were higher performing, as Table 1-3 shows. The state identified schools proportionally by level based on the eligible population of schools, with 38 elementary, 28 middle, and 12 high schools assigned to treatment.

Table 1-3. School sample characteristics

	NCT	Control
Urbanicity		
City	0.0 (0.11)	0.1 (0.23)
Suburb	0.0 (0.00)	0.1 (0.22)
Town	0.1 (0.25)	0.1 (0.30)
Rural	0.9 (0.27)	0.8 (0.41)
School level		
Elementary	48.7 (50.31)	57.3 (49.56)
Middle	35.9 (48.28)	33.6 (47.33)
High	15.4 (36.31)	9.1 (28.80)
Student achievement		
2015 performance composite (centered)	-5.1 (4.31)	9.8 (5.43)
EVAAS growth score	68.6 (10.36)	68.5 (10.64)
Teacher qualifications		
Percent novice teachers	32.5 (12.57)	26.8 (12.27)
Percent National Board Certification	7.7 (4.64)	11.5 (7.24)
Student demographics		
Minority percent	84.7 (12.44)	60.2 (21.70)
Economically disadvantaged	82.2 (12.12)	72.7 (14.66)
School characteristics		
Per pupil spending	10217.7 (2264.00)	9426.0 (1863.50)
Average Daily Membership	429.2 (172.67)	498.8 (226.64)

NOTE: Means and standard deviations on baseline measures based on 331 eligible schools.

Data and Measures

This analysis draws from a longitudinal database of statewide administrative data maintained by the University of North Carolina-Chapel Hill's Educational Policy Initiative at Carolina (EPIC). The database contains data on all students, teachers, principals, and schools in North Carolina public schools. Our analysis uses student-level data to estimate the effect of NCT on student achievement and teacher-level data to estimate the effect on teacher turnover.

Outcome measures

We estimate the effect of NCT on end-of-grade (EOG) and end-of-course (EOC) test scores. Students in North Carolina take math and reading EOGs each year in third through eighth grade, science EOGs in fifth and eighth grade, and EOCs in Math 1, English II, and Biology. Exams are administered in the final 10 instructional days of the school year for year-long courses and the final five instructional days of fall semester for half-year block EOC courses taken in the fall. We operationalize teacher turnover as leaving the school, either to move to another school or leave North Carolina public schools altogether. Teacher turnover is measured during and at the end of the school year, so a teacher who does not return to her 2015-16 school in the 2016-17 school year would be counted as having turned over in 2015-16.

Assignment variable

The state assigned schools to receive services based on the 2014-15 school performance composite, a measure that represents the EOG and EOC exam passage rate (abbreviated below as GLP, for grade-level proficiency). To account for differences in passage rates by exam and ensure the proportion of treated elementary, middle, and high schools roughly matched the eligible sample's proportion of schools at each level, the state set separate cutoffs for elementary, middle, and high schools. The cutoff was 31.1 for elementary schools, 33.8 for middle schools, and 26.0 for high schools. Schools below these thresholds were targeted for services. For the analysis, we center the performance composite at the threshold by school level.

Teacher effectiveness

To explore whether teacher mobility was intentional and strategic, we draw from two lagged measures of teacher effectiveness. Subject-specific value-added scores (Education Value-Added System, or EVAAS) provide a measure of teacher effectiveness for teachers of tested grades and subjects, while the teacher's evaluation ratings as measured by the North Carolina Educator Effectiveness System (NCEES) are available for teachers of tested and untested grades and subjects. We use EVAAS scores calculated from EOCs and EOGs, as well as mClass reading assessments in kindergarten through third grade. About one-third of teachers in the sample have lagged scores in each outcome year. Teachers receive one of three ratings based on their EVAAS score for a given subject: they *meet expected growth* if they are within 2 points of predicted growth on the EVAAS scale, *exceed expected growth* at more than 2 points above, and *do not meet expected growth* at more than 2 points below. We use these cutoffs to place teachers in effectiveness categories. Specifically, we code a teacher as "highly effective" if she has a lagged EVAAS score that exceeds expected growth, "low effectiveness" if she has a lagged EVAAS score that does not meet expected growth, and "mid effectiveness" if all EVAAS scores fall in the meets expected growth category.[3]

[3] About 26% of teachers with lagged EVAAS scores are low EVAAS, 63% are mid, and 11% are high.

NCEES includes five standards: (1) teacher leadership, (2) establishing a respectful learning environment for diverse students, (3) content knowledge, (4) facilitate learning for students, and (5) reflecting on practice. Teachers receive ratings of 1 to 5 on each rating, with 1 being the lowest rating a teacher can receive and 5 the highest. Because teachers with more than three years of experience are only required to be evaluated on standards 1 and 4, we draw the NCEES measures from these two standards. We observe lagged NCEES ratings on each of these standards for about 70 percent of the sample during the outcome years. We generate two different NCEES effectiveness measures—one for standard 1 and one for standard 4. The modal rating in the sample on both measures is a 3. We again place teachers into three effectiveness categories based on these lagged NCEES ratings: "low effectiveness" for teachers with a 1 or 2, "mid effectiveness" for teachers with a 3, and "highly effective" for teachers with a 4 or 5.[4]

Using EVAAS and NCEES, we end up with three categorical measures of teacher effectiveness: high, mid, and low EVAAS; high, mid, and low NCEES standard 1; and high, mid, and low NCEES standard 4. Each has distinct advantages and disadvantages. EVAAS contains the most variation but restricts the sample to just teachers who were in tested grades and subjects the prior year. NCEES captures more of the sample but classifies very few teachers in the low category (about 2% of teachers in the sample).

Implementation

We examine three implementation measures to determine whether variation in implementation was associated with differences in outcomes—focusing specifically on dimensions of implementation with substantial variation across schools. These three dimensions

[4] On NCEES standard 1, about 49% of teachers with lagged scores in the sample are high, 49% are mid, and 2% are low. On standard 4, about 41% are high, 57% are mid, and 2% are low.

are the timing of the CNA, the presence of a CNA unpacking, and the dosage of coaching. We collected measures for each of these variables directly from the state and validated them through site visits to treatment schools and phone interviews with control school principals. The state did not provide coaching to control schools, and, according to documents provided by the state, did not undertake comprehensive needs assessments in control schools.

CNA timing. We drew from the state's CNA and unpacking calendar to categorize schools by CNA timing. The state prioritized scheduling CNAs in schools that had not received one in the three years prior to the beginning of supports. We therefore placed schools in four categories according to CNA timing: (1) CNA was more than three years before services began or did not occur at all—these are schools that did not receive the CNA component of the intervention on the timeline set by the theory of change; (2) CNA was during the three years before the start of services—these are schools for which the CNA was implemented as specified by the theory of change timeline but for which CNA findings may have no longer been relevant because of staff turnover or changing school dynamics; (3) CNA was in spring 2016—these are schools that received a CNA during the intervention but during the school year and after they would have already developed and begun to implement their school improvement plans; and (4) CNA was during the 2016-17 school year, the second year of supports. Because all services were intended to build from the CNA, we hypothesized that schools not receiving CNAs or not receiving them within a useful time period for planning might suffer from less coherent services or potentially lead to a disruption of school improvement efforts already in progress under the SIP developed at the beginning of the school year.

CNA unpacking. Also drawing from the state CNA and unpacking schedule, we created a single dichotomous measure denoting whether or not the school received an unpacking.[5] Forty-nine schools received an unpacking and 26 did not. Because the unpacking was intended to build from the CNA and provide schools with a path forward for the school improvement plan, we hypothesized that schools that received unpackings may have been able to develop more targeted improvement plans leading to better outcomes.

Coaching dosage. Using coaching reports provided by the state, we counted the number of school transformation, instructional, and total coaching visits carried out during the intervention period. The dosage measure takes three values: high dosage schools (top quartile), mid-dosage schools (middle 50%), and low dosage schools (bottom quartile). In this case, we hypothesized that schools receiving a lower dosage of services may have experienced negative effects if the amount of coaching received fell short of expectations and frustrated rather than building the capacity of principals and teachers.

Covariates

School-level variables include minority percentage, economically disadvantaged percentage, per pupil expenditures (PPE) and PPE squared, enrollment (average daily membership, or ADM) and ADM squared, and school level with elementary as the reference category. Teacher-level variables include female and race with white as the reference category. Student-level variables include female, race with white as the reference category, disabled, academically gifted, limited English proficient, over-age for grade, and nonstructural transfer in. We define disabled as a current designation with any exceptionality code other than

[5] We did not categorize schools by unpacking timing as we did by CNA timing because unpacking timing overlapped closely with CNA timing. Findings using unpacking timing are similar to those using CNA timing.

academically gifted. We define over-age as having a birthdate that would place the student in a grade level above the grade level assigned. We define nonstructural transfer in as a transfer that occurs into the observed school prior to the maximum grade of the prior school (e.g., transferring into the observed school in 7th grade when the student's prior school went through 8th grade).

Empirical Strategy

Main effects

We estimate the effect of NCT using a regression discontinuity design that exploits the jump in probability of assignment to treatment at the cutoff (Imbens & Lemieux, 2007). We begin with an intent-to-treat (ITT) estimate that takes the form

$$y_{is} = \beta_0 + \beta_1 I(GLP < 0)_s + \beta_2 f(GLP)_s + \beta_3 I(GLP < 0)_s \times f(GLP)_s + \gamma \boldsymbol{S}_s' + \sigma \boldsymbol{K}_i' + \varepsilon_{is} , \quad (1)$$

where *y* is the outcome for student or teacher *i* in school *s*, *GLP* represents the forcing variable, *I(GLP)* is an indicator for treatment eligibility that takes a value of 1 in schools below the assignment threshold, *f(GLP)* is a flexible function of the distance from the cutoff, the interaction between the treatment eligibility variable and forcing variable allows for a different slope on either side of the cutoff, and ε is an idiosyncratic error term clustered at the school level. In a second set of models, we add vectors of school-level covariates, ***S'***, and individual-level covariates, ***K'***, to increase precision. The individual-level covariates are student level in models predicting student test score and teacher level in the teacher turnover models. We also include the student's lagged test score on the right-hand side of the student achievement model. β_1 is the coefficient of interest, representing the estimated discontinuity at the cutoff. To model the effect of NCT around the cutoff, we estimate locally weighted linear regressions using a triangular kernel within the bandwidth calculated using the mean square error (MSE)-optimal bandwidth selection procedure described by Calonico, Cattaneo, & Titiunik (2014), which accounts for the clustered assignment of schools to treatment.[6]

This ITT analysis is the policy-relevant estimator because it represents the estimated effect of assignment to treatment. However, while eligibility for treatment was a strong predictor of receiving treatment, noncompliance occurred in schools above and below the cutoff. We therefore estimate a treatment on the treated (TOT) estimate using a two-stage least squares (2SLS) model in which we instrument NCT with treatment eligibility. The first stage of the 2SLS model takes the form

$$NCT = \alpha_0 + \alpha_1 I(GLP < 0)_s + \alpha_2 f(GLP)_s + \alpha_3 I(GLP < 0)_s \times f(GLP)_s + \gamma \boldsymbol{S}_s' + \sigma \boldsymbol{K}_i' + u_{is}\ , \tag{2}$$

where being in turnaround status (NCT) is a function of a treatment eligibility indicator, *I(GLP≤0)*, that takes a value of 1 if the school was below the treatment threshold; a flexible function of the distance from the cutoff, *f(GLP)*; and an interaction between the two. In the set of models with covariates, we include the vectors of school- and individual-level covariates in the first stage as well. We then estimate the second stage as

$$y_{is} = \beta_0 + \beta_1 (\widehat{NCT})_s + \beta_2 f(GLP)_s + \beta_3 I(GLP < 0)_s \times f(GLP)_s + \pi \boldsymbol{S}_s' + \rho \boldsymbol{K}_i' + \varepsilon_{is}\ , \tag{3}$$

where the predicted outcome, *y*, for student or teacher *i*, is a function of the predicted *NCT* indicator, and the model then follows the same format as the first stage. This approach allows us to estimate treatment effects using the schools that complied with their treatment assignment, with β_1 providing an estimated local complier-adjusted treatment effect. The fuzzy RD is our

[6] We use the rdrobust package in Stata to estimate the optimal bandwidths and the RD models (Calonico et al., 2017).

preferred model because it accounts for noncompliance and reflects the estimated treatment effect for compliers.

The TOT estimates would be biased if the instrument failed to meet the exclusion restriction, which requires that the instrument affects the outcome only through the instrumented variable (Angrist et al., 1996). In other words, having a performance composite below the threshold needs to affect student and teacher outcomes only through its effect on the likelihood of receiving turnaround services. The ITT estimates would not be subject to the same bias. While we cannot strictly test whether the exclusion restriction is met, we did survey principals in treatment and control schools to examine implementation in treatment schools and whether control schools may have received similar services. We draw from these data to descriptively examine whether (a) non-takers of assignment to treatment report similar levels of coaching to comparison group compliers, and (b) always-takers report similar levels of coaching to treatment group compliers. Using this descriptive analysis, we do not find differences between these groups in the probability of reporting receipt of school transformation or instructional coaching (Table A-1-1 provides crosstabulations of these groups with chi-square tests). While this approach to examining the exclusion restriction is limited to the schools for which we have survey responses on the relevant questions, we believe that these findings, combined with the similarity of the ITT and TOT estimates, suggest that any violation of the exclusion restriction that we are unable to observe would likely be minor and have only a negligible effect on the TOT estimates.

We stack all subjects in our main student achievement specification but also include separate models for math, reading, and science in the appendix. The lagged test score on the right-hand side of the equation is from one year prior for fourth- through eighth-grade math and reading. For high schools, the lag is from the eighth-grade EOG exam, which is two years prior for reading and most often one year prior for math. In science, there are two to three years between the lagged score and the outcome score.[7] Because the teacher turnover outcome is a binary indicator for whether the teacher turned over in a given school year, the teacher turnover models are linear probability models in which the RD estimate can be interpreted as the difference in probability of turnover associated with being in a treatment school relative to a control school at the cutoff.

We also estimate the model within a series of alternative bandwidths, including 50% and 200% of the CCT bandwidth, the optimal bandwidth proposed by Imbens & Kalyanaraman (IK, 2009), 200% of the IK bandwidth,[8] and finally on the full sample of treatment and control schools for which we have implementation data. We cluster standard errors at the school level.[9] Because coaching did not begin until spring 2016—i.e., the second semester of the intervention—we measure the outcomes separately for each year of treatment. The 2016 estimate represents the effect of a single semester of coaching in all schools and a CNA in most schools, while the 2017 estimate represents the effect of a full year of coaching services.[10]

Teacher effectiveness

After estimating the effects of the intervention on student achievement and teacher turnover, we conduct an additional analysis to examine the effectiveness of teachers who left the

[7] Because lagged test scores vary by subject area and grade level, we also estimate models without the lagged test score and find similar results.

[8] We do not estimate on 50% of the IK bandwidth because the bandwidth size—which unlike the CCT procedure does not account for the clustering of students within schools—includes only three schools above the cutoff.

[9] We also estimate the same set of test score models clustering standard errors at the student level to account for clustering of students across multiple exams in a year. However, the standard errors clustered at the student level are smaller, so the estimates with standard errors clustered at the school level that we show represent a more conservative approach.

[10] Because we include the lagged test score on the right side of the model, the estimated effect on student achievement in 2017 represents the effect of NCT in the second year of services after partialing out any effect from the first year.

schools and those who entered. Specifically, we are interested in whether the effects of NCT on teacher turnover and new-to-school teachers differ between more and less effective teachers. We define three treatment groups—high, mid, and low effectiveness teachers in NCT schools—using EVAAS scores, NCEES standard 1, and NCEES standard 4. To estimate the effects of NCT on each of these three groups of teachers, we implement analyses following the fuzzy RD framework with equations that are analogous to equations (2) and (3) predicting two dichotomous outcomes—turnover and being new to school—using three separate treatment groups rather than a single treatment. Specifically, to test for heterogeneity between teachers in each of three levels of effectiveness, we interact assignment to treatment with each of the three categories of teacher effectiveness. In order to estimate within-group differences between NCT and control schools, we also include indicators for high and low-effectiveness teachers. Because we have three treatment groups, we estimate three first-stage models predicting turnaround status within each of the three groups based on teacher effectiveness category. Equation 4 represents the equation for the highly effective group of teachers. In the other two first-stage equations we substitute *HighlyEffective* with *MidEffectiveness* and *LowEffectiveness* indicators. Otherwise, all three first stage equations are identically specified.

$$\begin{aligned}\Pr(NCT_s|HighlyEffective_i) &= \alpha_0 + \alpha_1(HighlyEffective_i \times I(GLP < 0)_s) + \alpha_2 f(GLP)_s \\ &+ \alpha_3 I(GLP < 0)_s \times f(GLP)_s + \alpha_4 HighlyEffective + \varepsilon_{is}\end{aligned} \quad (4)$$

The first-stage outcome in equation 4 is the predicted probability of being in a treated school for highly effective teachers. In other words, the first stage estimates the probability of being in a treated school, conditional on the school's assignment to treatment and the teacher being in the highly effective category. The coefficient estimate represented by α_1 provides the estimated effect of a teacher being in a particular effectiveness group in a school below the cutoff on the probability of treatment. The first-stage equations produce three separate predicted variables to carry into the second stage—one for each of the three treatment groups represented by teachers of high, mid, or low effectiveness. We then include the fitted values of the dependent variables from the three first-stage equations as predictors in the second stage:

$$\begin{aligned} y_{is} = \beta_0 &+ \beta_1(NCT_s \mid HighlyEffective_i) + \beta_2(NCT_s \mid MidEffectiveness_i) \\ &+ \beta_3(NCT_s \mid LowEffectiveness_i) + \beta_4 f(GLP)_s + \beta_5 I(GLP < 0)_s \\ &\times f(GLP)_s + \beta_6 HighlyEffective_i + \beta_7 LowEffectiveness_i + \varepsilon_{is} \end{aligned} \tag{5}$$

The outcome (turnover or new to school, represented as *y*) for teacher *i* in school *s* is estimated using the same approach as equation 3, but in equation 5, the three teacher effectiveness predicted values allow separate within-effectiveness-group estimates of the probability of turnover or being new to school. The $NCT_s \mid HighlyEffective_i$ variable is the predicted value of the dependent variable from Equation 4 above, so β_1 represents the complier-adjusted local average treatment effect for highly effective teachers in NCT schools. The $NCT_s \mid MidEffectiveness_i$ and $NCT_s \mid LowEffectiveness_i$ variables represent the predicted values of the dependent variables from the two parallel first-stage equations. We present estimates from these models without additional covariates, though the estimates are robust to inclusion of school- and teacher-level covariates.

Evidence of strategic staffing would be apparent when examining the β_1 and β_2 coefficient estimates. In the model estimating effects on turnover, a negative and significant estimate on the highly effective group (β_1) would provide evidence that treatment schools retained more effective teachers than control schools, while a positive and significant estimate on the low effectiveness β_3 would provide evidence that more of the less effective teachers left treatment schools than the control group schools. In the model predicting new-to-school teachers, a positive and significant estimate on the highly effective group (β_1) would provide evidence that treatment schools hired more effective teachers, while a negative and significant estimate on low effectiveness group (β_3) would provide evidence that treatment schools hired fewer ineffective teachers relative to control schools.

Implementation

We use a similar approach to test for heterogeneous effects by each of the three dimensions of implementation, replacing the teacher effectiveness group with the appropriate implementation category (by CNA timing group; whether or not the school received an unpacking; and high, mid, and low coaching dosage) and the outcome with student achievement. While equations (4) and (5) represent a traditional moderation approach comparing groups of teachers in treatment schools with similarly effective teachers in control schools, the implementation analysis simply compares the performance of students in groups of treatment schools with the performance of students in all control schools within the bandwidth. We take this approach because control schools did not have CNAs scheduled during the study period and therefore could not be placed into subgroups based on CNA timing. We illustrate the empirical approach with one of the three dimensions of implementation that we examined, the timing of the CNA. The first-stage model for the group of schools that did not receive CNAs or received one prior to 2014 therefore takes the form

$$\begin{aligned}\Pr(NCT|NoCNA)_s &= \alpha_0 + \alpha_1\big(NoCNA \times I(GLP < 0)\big)_s \\ &+ \alpha_2\big(2014or2015CNA \times I(GLP < 0)\big)_s \\ &+ \alpha_3\big(Spring2016 \times I(GLP < 0)\big)_s + \alpha_4\big(201617 \times I(GLP < 0)\big)_s \\ &+ \alpha_5 f(GLP)_s + \alpha_6 I(GLP < 0)_s \times f(GLP)_s + \alpha_7 Score_{it-1} + \varepsilon_{is}\end{aligned} \tag{6}$$

where the first-stage outcome is the predicted probability of being in a treated school that did not receive a CNA or received one prior to 2014. Here, we estimate separate first-stage equations for each implementation group; in other words, for the CNA timing analysis, we estimate three additional first-stage equations predicting the probability of being in a treated school that received a CNA in 2014 or 2015, being in a treated school that received a CNA in spring 2016, and being in a treated school that received a CNA in the 2016-17 school year. We carry the predicted values from each of the four first-stage equations into the second stage:

$$
\begin{aligned}
Score_{is} = \beta_0 &+ \beta_1(NCT|NoCNA)_s + \beta_2(NCT|2014or2015CNA)_s \\
&+ \beta_3(NCT|Spring2016CNA)_s + \beta_4(NCT|2016-17CNA)_s \\
&+ \beta_5 f(GLP)_s + \beta_6 I(GLP < 0)_s \times f(GLP)_s + \beta_3 Score_{it-1} + \varepsilon_{is}
\end{aligned} \tag{7}
$$

where the test score for student *i* in school *s* is a function of each of the predicted probabilities of being in the four categories of treatment by CNA timing from the first-stage equations, the forcing variable, an interaction between the forcing variable and being assigned to treatment, the student's lagged test score, and vectors of school and student covariates. The coefficient estimates on the four separate treatments represent the estimated effect of being in an NCT school that received a CNA in a particular time period on student achievement, relative to students in all control schools at the cutoff. Estimates on β_1 through β_4 that are different from one another would provide evidence that variation in implementation was associated with differences in outcomes.

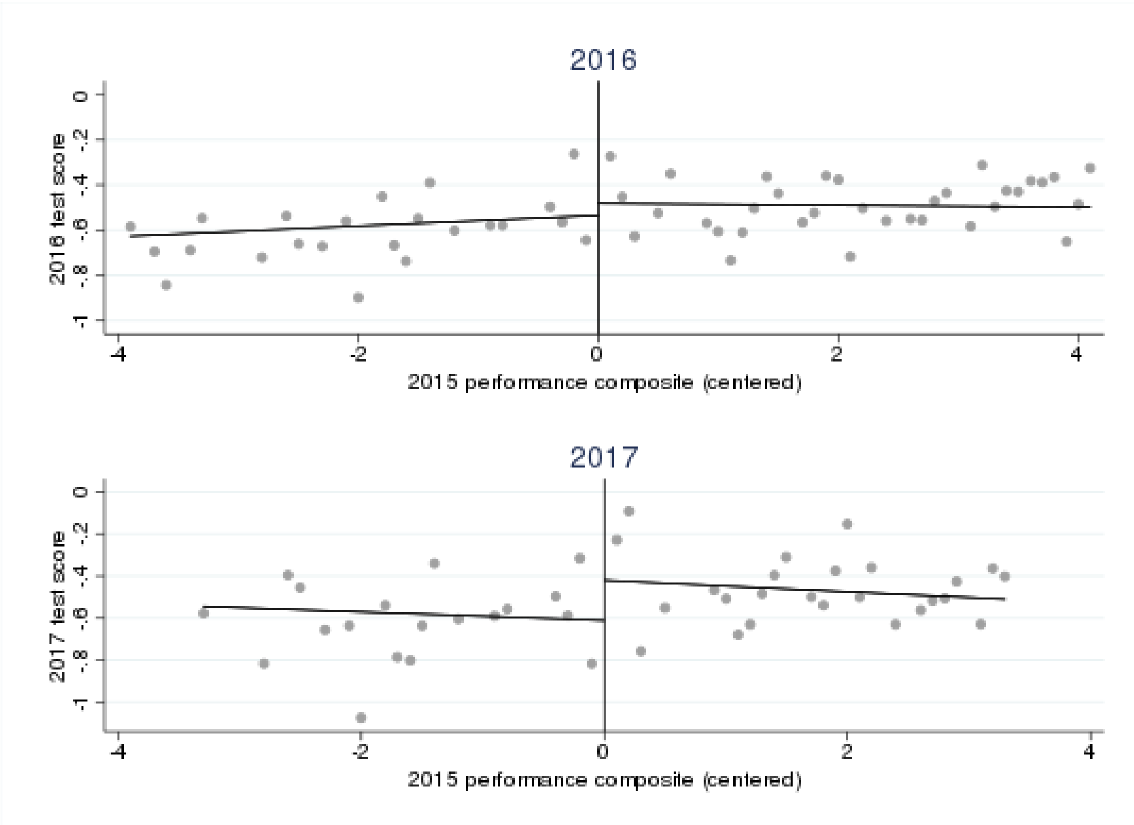
We consider the implementation analyses correlational rather than causal because schools were not randomly assigned to implementation variation such as CNA receipt in a particular time period. We do examine whether the four subgroups of schools that vary with CNA timing appear to have different effects at the time of assignment to treatment. To do so, we regress the school baseline performance composite on implementation groupings. We do not find that the CNA timing groups are significant predictors of baseline performance. These results, along with results for other implementation groupings, are provided in Table A-1-2.[11]

Results

We find consistent evidence that NCT had a negative effect on student achievement in 2017 and neither a positive or negative effect in 2016. Figure 1-2 provides a graphical representation of these results within the preferred bandwidth. The vertical distance between the fit lines on either side of the cutoff represents the difference in outcomes associated with being in a school assigned to treatment. The 2017 panel provides graphical evidence of a decrease in student achievement among schools below the cutoff in the second year of services.

[11] While we do not find significant effects on our primary implementation analysis, which is focused on CNA timing, we do find significant effects in two implementation groupings that we show in the appendix. First, having an unpacking in 2014 or 2015 was associated with a lower baseline performance composite than having no unpacking. Second, schools in the highest quartile of instructional coaching visit dosage had lower predicted baseline performance composites than schools in the middle 50% of instructional coaching visit dosage.

Figure 1-2. Student achievement by distance from assignment threshold



NOTE: Markers represent bin averages within CCT bandwidths and lines are linear fit. Estimation using triangular kernel within preferred CCT bandwidth, with average bin width of .006 to left of cutoff and .007 to right of cutoff in 2016, and .007 to left of cutoff and .010 to right of cutoff in 2017.

Table 1-4 displays the ITT estimates separately for 2016 (Panel A) and 2017 (Panel B). Model 1, which estimates within the preferred CCT bandwidth, shows that assignment to treatment has a negative effect, -0.12 standard deviations, on test scores in the second year of treatment. This result is robust to alternative bandwidths (Models 3–6) and inclusion of covariates (Models 2, 4, and 6).

Table 1-4. ITT estimates *(outcome=test score)*

Panel A: 2016

	(1)	(2)	(3)	(4)	(5)	(6)
	CCT		50% CCT		200% CCT	
ITT	-0.063	-0.034	-0.130*	-0.090**	-0.024	-0.020
	(0.0581)	(0.0443)	(0.0583)	(0.0295)	(0.0403)	(0.0356)
Covariates		X		X		X
Bandwidth	4.1	4.1	2.1	2.1	8.3	8.3
N	195437	195437	195437	195437	195437	195437
N Bandwidth	50731	50731	23415	23415	92514	92514
T schools in BW	36	36	22	22	66	66
C schools in BW	51	51	19	19	102	102

Panel B: 2017

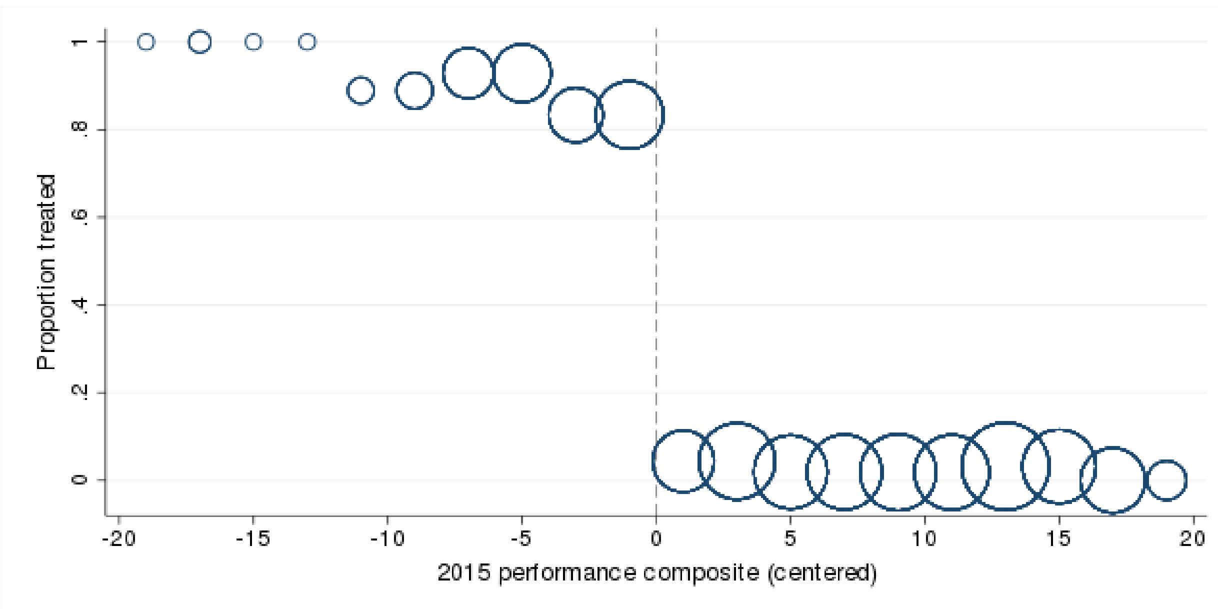
	(1)	(2)	(3)	(4)	(5)	(6)
	CCT		50% CCT		200% CCT	
ITT	-0.123*	-0.131**	-0.172**	-0.221***	-0.101*	-0.093*
	(0.0521)	(0.0403)	(0.0535)	(0.0249)	(0.0395)	(0.0365)
Covariates		X		X		X
Bandwidth	3.3	3.3	1.7	1.7	6.7	6.7
N	195099	195099	195099	195099	195099	195099
N Bandwidth	39423	39423	18624	18624	77420	77420
T schools in BW	31	31	18	18	55	55
C schools in BW	37	37	13	13	84	84

Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All models include lagged score and subject fixed effects on the right side, with math as the reference category.

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

These ITT models provide the policy-relevant estimator, but do not account for noncompliance with treatment assignment, which occurred on both sides of the cutoff. The probability of treatment is high for schools assigned to treatment and low for those not assigned to treatment, but Figure 1-3 shows that a small proportion of schools below the cutoff did not receive treatment and a small proportion of schools above the cutoff did receive treatment. The fuzzy RD accounts for this noncompliance by providing the estimated local average treatment effect of NCT for compliers.

Figure 1-3. Proportion treated by forcing variable



NOTE: Markers represent bin averages. Bin width is 2. Marker sizes weighted by number of schools in bin.

The TOT estimates from the fuzzy RD are provided in Table 1-5. These complier-adjusted treatment effects are similar to the ITT estimates, with an estimated effect of -0.13 in 2017 in our preferred model. The similarity in terms of both magnitude and significance of the ITT and TOT estimates suggest that the noncompliance has a negligible effect on the ITT estimates. We proceed by showing TOT estimates from the fuzzy RDs in the remainder of the manuscript.

Similar to the ITT estimates, the 2017 TOT estimates displayed in Table 1-5 are consistently negative and significant across the three bandwidths and with and without covariates. In the analytical sample defined by the narrowest bandwidth, we find a negative effect of NCT in 2016. This pattern of effects is similar when we estimate within alternative bandwidths and using the full sample. The effects in 2017 are consistently negative and significant and the effects in 2016 are only significant when the analytical sample is defined by narrower bandwidths calculated using the bandwidth selection procedure described in Imbens & Kalyanaraman (2009). [12]

[12] The IK bandwidths are narrower than the CCT bandwidths. The estimate in the 17 schools within the IK bandwidth is -.186 and the estimate in the 35 schools in the 200% IK bandwidth is -.146.

Table 1-5. TOT estimates *(outcome=test scores)*

Panel A: 2016

	(1) CCT	(2)	(3) 50% CCT	(4)	(5) 200% CCT	(6)
TOT	-0.066	-0.039	-0.148**	-0.135***	-0.027	-0.023
	(0.0592)	(0.0512)	(0.0511)	(0.0389)	(0.0449)	(0.0407)
Covariates		X		X		X
Bandwidth	4.1	4.1	2.1	2.1	8.3	8.3
First-stage *F*-stat	120.78	113.64	35.28	36.84	213.16	213.16
N	195437	195437	195437	195437	195437	195437
N Bandwidth	50731	50731	23415	23415	92514	92514
T schools in BW	36	36	22	22	66	66
C schools in BW	51	51	19	19	102	102

Panel B: 2017

	(1) CCT	(2)	(3) 50% CCT	(4)	(5) 200% CCT	(6)
TOT	-0.131*	-0.170**	-0.198***	-0.420***	-0.111*	-0.109*
	(0.0517)	(0.0541)	(0.0560)	(0.0710)	(0.0433)	(0.0433)
Covariates		X		X		X
Bandwidth	3.3	3.3	1.7	1.7	6.7	6.7
First-stage *F*-stat	81.72	88.74	29.81	49.14	222.01	211.41
N	195099	195099	195099	195099	195099	195099
N Bandwidth	39423	39423	18624	18624	77420	77420
T schools in BW	31	31	18	18	55	55
C schools in BW	37	37	13	13	84	84

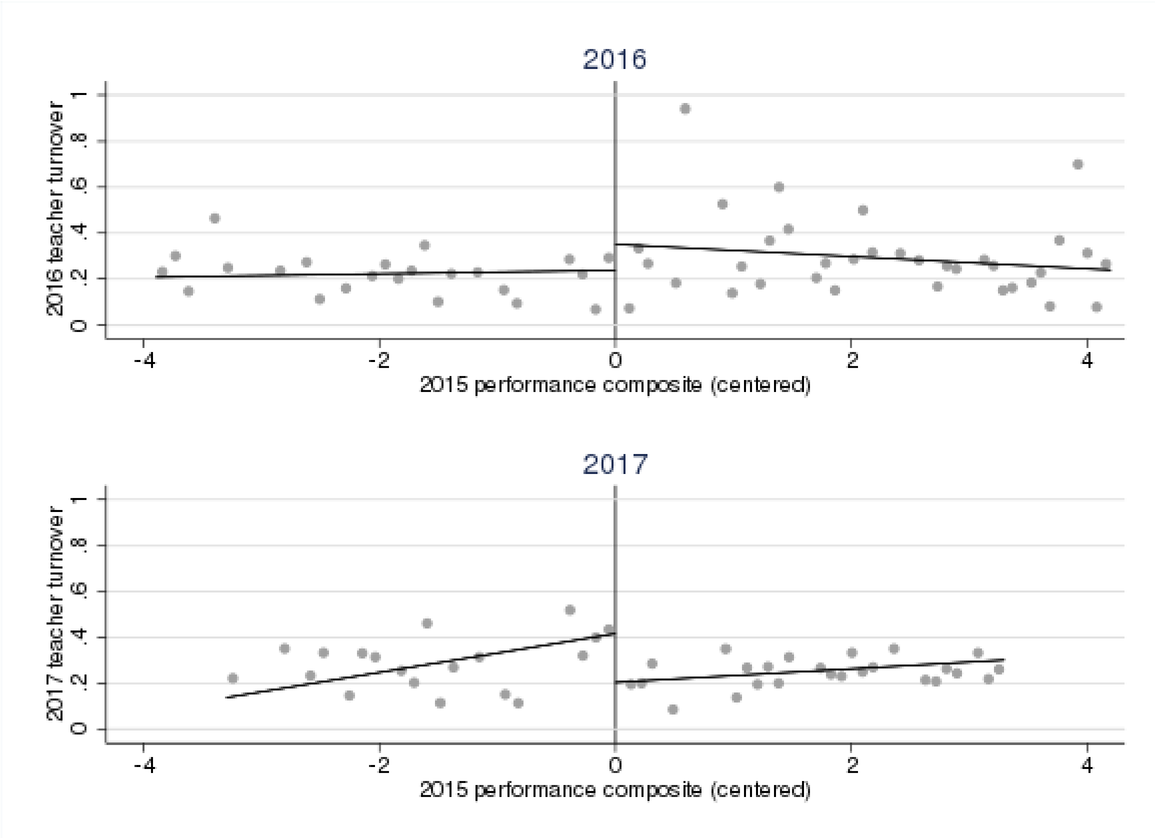
Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All models include lagged score and subject fixed effects on the right side, with math as the reference category.

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

Central to the validity of our estimates is the ability to rule out a weak instrument (Stock & Yogo, 2002). The recommended minimum first-stage *F*-statistic on the treatment indicator to show that the instrument is a sufficiently strong predictor of treatment is 16 (What Works Clearinghouse, 2017). All first-stage *F*-statistics exceed this criterion as shown in Table 1-5.

The results are qualitatively similar across subject areas, with consistently negative point estimates for math, reading, and science across all specifications in both years. The significant negative effects in 2017 appear to be driven by reading scores, where we estimate an effect of -0.16 standard deviations of student achievement (Table A-1-4). We also find qualitatively similar results when we estimate on test score levels rather than conditioning on lagged achievement values, shown in Table A-1-5, providing some evidence that the negative effects aren't driven by idiosyncrasies of the sample of students with lagged scores or the variation in timing for lagged score in high school and science exams. Finally, the negative effects of NCT appear to be consistent across all school levels, although we do not have a strong enough first stage to obtain valid TOT estimates in elementary schools (Table A-1-6).

Figure 1-4. Teacher turnover by distance from assignment threshold



NOTE: Graph based on school-level averages of dichotomous teacher turnover variable. Markers represent individual school averages and lines are linear fit. Estimation using triangular kernel within preferred CCT bandwidth.

Teacher turnover. We also find evidence that teachers in NCT schools were more likely to turn over in 2017 but neither more nor less likely to turn over in 2016, displayed visually in Figure 1-4 and numerically in Table 1-6. In 2017, teachers in NCT schools were 22.5 percentage points more likely to turn over than control school teachers. These estimates are consistent across bandwidths and robust to the inclusion of covariates (Table A-1-7).[13]

[13] Results from the full analytical sample and alternative IK bandwidths are presented in in Tables A-7 and A-8. While a weak first stage in the 50 percent IK bandwidth for 2017 precludes valid inferences for the TOT estimate within this bandwidth, a sharp specification finds significant increases in teacher turnover in the narrowest bandwidth and across other bandwidths (Table A-1-8).

Table 1-6. TOT estimates *(outcome=teacher turnover)*

Panel A: 2016

	(1)	(2)	(3)	(4)	(5)	(6)
	CCT		50% CCT		200% CCT	
TOT	-0.044	-0.103	0.091	0.164	-0.074	-0.087
	(0.0945)	(0.0822)	(0.1316)	(0.1388)	(0.0600)	(0.0547)
Covariates		X		X		X
Bandwidth	4.1	4.1	2.1	2.1	8.3	8.3
First-stage *F*-stat	31.47	32.60	7.84	6.45	74.13	76.74
N	10770	10770	10770	10770	10770	10770
N Bandwidth	2658	2658	1240	1240	5270	5270
T schools in BW	35	35	21	21	64	64
C schools in BW	51	51	19	19	102	102

Panel B: 2017

	(1)	(2)	(3)	(4)	(5)	(6)
	CCT		50% CCT		200% CCT	
TOT	0.225^{**}	0.204^{*}	0.357^{**}	0.393	0.128	0.126^{*}
	(0.0819)	(0.0891)	(0.1342)	(0.2676)	(0.0669)	(0.0604)
Covariates		X		X		X
Bandwidth	3.3	3.3	1.7	1.7	6.7	6.7
First-stage *F*-stat	24.01	24.80	6.86	5.52	66.75	69.06
N	10492	10492	10492	10492	10492	10492
N Bandwidth	2078	2078	940	940	4280	4280
T schools in BW	30	30	17	17	53	53
C schools in BW	37	37	13	13	84	84

Estimates from linear probability models. Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. Red outlines denote first-stage F statistics on the treatment indicator smaller than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage.

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

Compositional effects of teacher turnover. While teacher turnover has been found to generally have negative effects (Hanushek, Rivkin, & Schiman, 2016; Henry & Redding, 2020; Ronfeldt, Loeb, & Wyckoff, 2013), strategically replacing lower performing teachers with more effective teachers can have positive effects—especially in very low performing schools (Adnot et al., 2017; Henry et al., 2020; Strunk et al., 2016; Zimmer et al., 2017). By extension, a negative compositional effect of teacher turnover may help to explain negative effects on student achievement. If turnover of effective teachers was particularly high in 2016, or if replacement teachers in 2017 were worse on average than departing teachers, these staffing changes could help explain the negative effects in 2017. Meanwhile, lower turnover of effective teachers or higher turnover of ineffective teachers in 2017 might suggest that schools are engaging in strategic staffing for the future and that the negative effects in 2017 may be temporary.

We do not find consistent evidence that negative compositional effects were likely to have produced the negative effects on student achievement in 2017. If these negative effects were driven by turnover of highly effective teachers paired with replacement by less effective teachers, Table 1-7 would show positive point estimates on both *TOT x high effectiveness* in Panel A for 2016 (Columns 1-3) and *TOT x low effectiveness* in Panel B for 2017 (Columns 4-6). The former would suggest that highly effective teachers in treatment schools were more likely to turn over than their counterparts in control schools after the first year of services, while the latter would suggest that treatment schools were more likely than control schools to fill vacancies with less effective teachers. We do not detect significant effects on any of these coefficients. Similarly, the estimates on *TOT x high effectiveness* in Panel A for 2016 suggest highly effective teachers were no less likely to turn over in treatment than in control schools. To that end, we do not find evidence that the negative compositional effect of turnover drove negative effects on student achievement.

Table 1-7. TOT estimates on teacher turnover and new-to-school teachers by lagged teacher effectiveness

Panel A: Teacher turnover

	2016			2017		
	(1)	(2)	(3)	(4)	(5)	(6)
	Standard 1	Standard 4	EVAAS	Standard 1	Standard 4	EVAAS
Low effectiveness	-0.092	0.211	0.078	-0.338	0.083	0.060
	(0.1278)	(0.1309)	(0.0615)	(0.4532)	(0.1457)	(0.0533)
High effectiveness	-0.048	-0.036	-0.088	-0.029	0.002	-0.022
	(0.0376)	(0.0540)	(0.0742)	(0.0231)	(0.0281)	(0.1021)
TOT x low effectiveness	0.082	-0.205	-0.142	1.023	0.558^{*}	0.157
	(0.1823)	(0.1989)	(0.1215)	(0.6286)	(0.2722)	(0.1185)
TOT x mid effectiveness	-0.057	-0.045	-0.017	0.221^{*}	0.193^{*}	0.137
	(0.0911)	(0.0880)	(0.1242)	(0.0952)	(0.0879)	(0.1059)
TOT x high effectiveness	0.025	0.055	0.004	0.155	0.186	0.104
	(0.1142)	(0.1338)	(0.1803)	(0.0820)	(0.1006)	(0.1296)
Constant	0.295^{***}	0.274^{***}	0.261^{**}	0.166^{***}	0.154^{***}	0.188^{**}
	(0.0841)	(0.0825)	(0.1003)	(0.0395)	(0.0382)	(0.0594)
N	1997	1997	1102	1568	1568	786

Panel B: New-to-school teachers

	2016			2017		
	(1)	(2)	(3)	(4)	(5)	(6)
	Standard 1	Standard 4	EVAAS	Standard 1	Standard 4	EVAAS
Low effectiveness	0.241*	-0.067***	0.032	0.307	0.069	-0.007
	(0.1223)	(0.0124)	(0.0383)	(0.2315)	(0.0915)	(0.0431)
High effectiveness	-0.036**	-0.054***	0.057	0.005	-0.019	0.026
	(0.0122)	(0.0117)	(0.0667)	(0.0137)	(0.0140)	(0.0319)
TOT x low effectiveness	-0.150	0.148	0.021	-0.375	-0.134	0.099
	(0.1624)	(0.0837)	(0.0641)	(0.2769)	(0.1336)	(0.0535)
TOT x mid effectiveness	-0.010	-0.015	0.094	0.010	-0.008	0.098
	(0.0257)	(0.0247)	(0.0544)	(0.0168)	(0.0166)	(0.0750)
TOT x high effectiveness	0.016	0.019	0.000	-0.005	0.027	0.029
	(0.0260)	(0.0208)	(0.1295)	(0.0213)	(0.0237)	(0.0818)
Constant	0.055**	0.062***	0.043	0.022	0.031*	0.035
	(0.0198)	(0.0173)	(0.0270)	(0.0128)	(0.0134)	(0.0527)
N	1997	1997	1102	1568	1568	786

NOTE: Effectiveness based on prior year NCEES (Columns 1-2 and 4-5) and EVAAS (Columns 3 and 6). NCEES standard 1 is teacher leadership. NCEES standard 4 is facilitating student learning. Low NCEES is defined as a score of 1 or 2 on 5-point scale, mid NCEES defined as score of 3, and high NCEES defined as 4 or 5. Low EVAAS is defined as an EVAAS score of <-2, which the state categorizes as not meeting expected growth, average EVAAS is defined as a score between -2 and 2, which the state categorizes as meeting expected growth, and high EVAAS is defined as an EVAAS score of >2, which the state categorizes as exceeding expected growth.

Standard errors clustered at the school level. All models estimated within CCT bandwidths calculated using the fuzzy test score models.

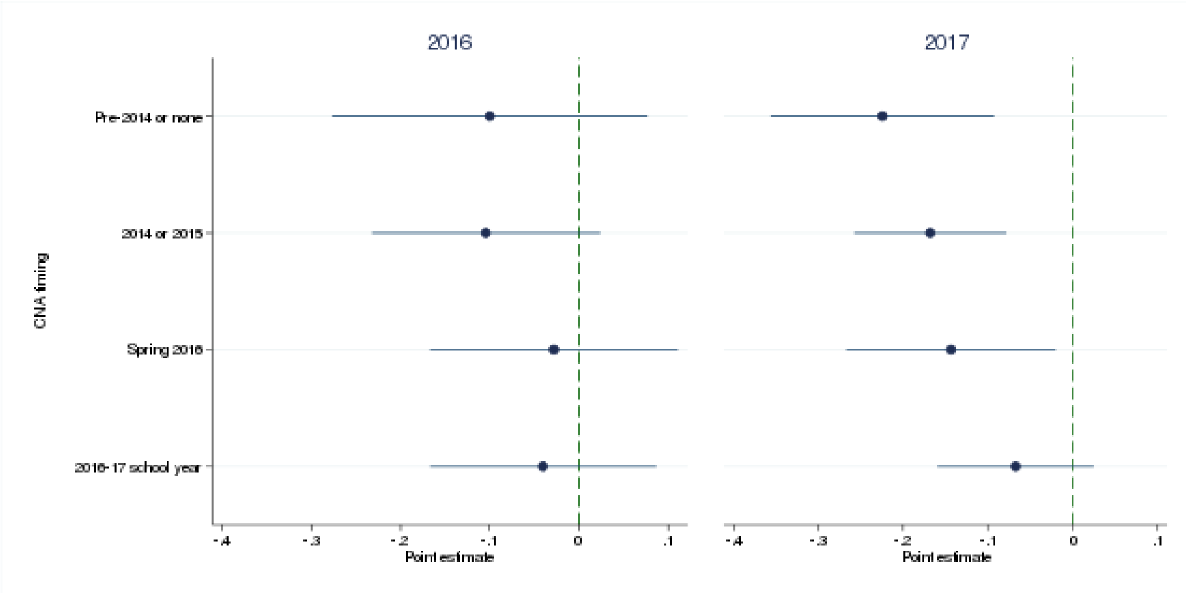
All first-stage F-statistics are greater than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage, except for the test statistic for TOT x average EVAAS in Model 6, which is 10.24.

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

Meanwhile, if the high turnover in 2017 were strategic, with treatment schools intentionally dismissing or coaching out their least effective teachers, we would observe positive estimates on *TOT x low effectiveness* in Panel A for 2017 (Columns 4-6). Significant positive effects for this group would provide evidence that the least effective teachers were more likely to turn over than their counterparts in control schools, suggesting the negative effects might be temporary as the reform schools re-staff. We do find that these estimates are descriptively positive and significant on one measure, but we also see that NCT teachers in all three effectiveness categories were descriptively more likely to turn over in 2017 than their counterparts in control schools. Taken together, these findings suggest teacher mobility in treatment schools was neither detrimental enough in 2016 to explain student achievement losses in the following year, nor was it clearly strategic in 2017 to augur future growth. Still, we cannot completely rule out either of these hypotheses given the relatively imprecise estimates in some of these models.

Implementation. While we do not find evidence of dosage effects or the availability of an unpacking after the CNA (Figure A-1-2 and Figure A-1-3), we do find suggestive evidence that the negative effects on student achievement in 2017 appear to be concentrated in three categories of CNA timing. In particular, Figure 1-5 shows that the negative effects in 2017 occurred in schools that did not receive CNAs at all, received CNAs in 2014 or 2015 before the intervention began, or received CNAs in spring 2016 while school improvement plans were being implemented and when coaching services were being delivered. We observe null effects in the 17 schools that received CNAs in the 2016-17 school year. While we believe the pattern of effects is informative, this finding should nevertheless be interpreted with caution given the overlapping confidence intervals for the four subgroup effects.

Figure 1-5. Heterogeneity of effects by Comprehensive Needs Assessment timing



NOTE: Estimates from fuzzy RD models with triangular kernel and 4 different treatments within preferred CCT bandwidths. Markers represent point estimates and spikes represent 95% confidence intervals. CCT bandwidths calculated using main fuzzy test score models. All first-stage *F*-statistics are greater than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage. Corresponding point estimates provided in Table A-1-11.

Qualitative data collected as a part of the overall evaluation provides some context for interpreting these results. Descriptively, the largest negative effects appear in schools that did not receive a CNA within a period useful for school improvement planning (more than two years before the NCT services began, if ever). The intervention delivered in these schools effectively undermined the theory of change, which predicated the reform strategy on an in-depth assessment of school needs drawing from multiple forms of data, including instructional observations. Schools receiving CNAs in 2014 or 2015, prior to the implementation of NCT in 2016, also present negative effects. These schools received services based on findings from before they were designated as eligible for NCT and, in many schools, before much of the staff carrying out the school improvement plans, including the principals, were in place. To that end, the needs identified among these schools—such as instructional quality in specific subjects or grades that are observed as part of the CNA process—may have been outdated, and services aligned to these needs again may have been misaligned with the current needs of the school. Moreover, the principals and school improvement teams in these schools may have been unaware that the CNA was conducted or of the particular needs that were identified, and thus unable to take the findings into account during the school improvement planning.

Finally, schools that received CNAs in spring 2016, which experienced negative effects that were descriptively weaker than the latter two groups, may have struggled due to two factors. First, CNA findings communicated in the middle of the school year may have disrupted implementation of the school improvement plan that was prepared during the prior fall, undermining commitment to the plan when school staff were preparing for state testing. Second, data collected from teachers and principals in the schools receiving CNAs in spring 2016 suggested weak communication between state and school staff concerning the CNA timing and process. During this time period, state agency personnel communicated about the CNA with principals and expected principals to communicate with their staff. Principals and teachers in these schools shared that they felt intimidated by state personnel conducting the CNAs, many staff were surprised and upset when observers showed up in their classrooms without prior notice, and many were demoralized by the description of the schools' inadequacies presented in the CNA reports after they had committed substantial effort to implementing the improvement plan. The evaluation team shared these formative findings with NCT leadership and staff in the summer of 2016 and later qualitative data collection suggests program staff became much more proactive in their communication with the schools receiving CNAs, which corrected the communication issues that arose in spring 2016. During interviews, school staff reported that they viewed CNAs conducted during the 2016-17 school year more favorably and our findings show no negative effects among this group of schools.

Validity checks

Two assumptions are critical to the validity of the RD design. First, there should be no manipulation of the forcing variable or cutoff; in other words, there should be no evidence that the value of the performance composite or the eligibility threshold was changed to influence treatment assignment in schools near the cutoff. Second, the functional form of the relationship between the outcome and forcing variable must be correctly specified on both sides of the cutoff. Additional essential assumptions for the validity of the fuzzy RD design are that treatment eligibility is a sufficiently strong predictor of compliance with assignment to treatment and there is no clear violation of the exclusion restriction. In this section, we describe the above assumptions in detail and then provide evidence that the data meet additional assumptions relevant to the validity and consistency of our estimates.

As described in the Data section above, the state determined the cutoff value of the assignment variable after schools administered exams based on the number of schools that could be served by NCT. Manipulation by schools is therefore highly unlikely because schools did not know before the exam window the proficiency rate threshold for assignment to treatment. Additionally, graphical analysis shows no evidence of manipulation of the forcing variable around the cutoff.[14] A McCrary test fails to reject the null of that there is no discontinuity in the density of the forcing variable within the optimal CCT 2016 and 2017 bandwidths.[15]

The second core assumption for the validity of the local average treatment effect estimate is that the functional form is correctly specified on either side of the forcing variable. To meet this condition, we estimate separate local linear regressions within the CCT bandwidths on either side of the cutoff. Figure 1-2 and Figure 1-4 above provide visual evidence that the relationships

[14] We show the density of the forcing variable across the full sample of eligible schools in Figure A-1-1
[15] 2016 p=.2768; 2017 p=.1773

are linear within the preferred bandwidths for student achievement and teacher turnover, respectively. We also estimate effects within several alternative bandwidths, including 50 percent of the CCT bandwidth, 200 percent of the CCT bandwidth, the IK optimal bandwidth, and 200 percent of the IK bandwidth, and find that both outcomes are robust to most of these alternative bandwidths and on the full sample (Table A-1-3 and Table A-1-7).

The fuzzy RD design requires that eligibility is a sufficiently strong predictor of participation. Figure 1-3 above clearly shows schools below the cutoff had a high probability of receiving services while schools below the cutoff had a low probability of receiving treatment. First-stage test statistics on the treatment eligibility indicator provide formal evidence that the forcing variable is a sufficiently strong predictor of participation. All first-stage F-statistics on the treatment indicator are above the minimum recommended threshold of 16 (What Works Clearinghouse, 2017) in our preferred models as described in the Results section above. The first stage does not meet suggested criteria for narrower alternative bandwidths in the teacher turnover models or for the elementary school models. We denote models with weak first stages using a red box around the test statistic.

Another key assumption for the RD estimates to be consistent is that relationship between the forcing variable and outcome would be smooth in the absence of the intervention. While we cannot test this condition directly because we cannot observe the outcomes for treatment schools in the absence of treatment, we provide evidence for the smoothness condition in two ways. First, we show that the treatment sample is within the recommended .25 standard deviation units of the control sample on key covariates associated with school performance, conditional on the forcing variable, within the 2016 and 2017 preferred CCT bandwidths. Table 1-8 shows effect sizes from a series of models estimating the baseline (2015) covariate value using the forcing variable and a triangular kernel within the preferred bandwidth for each year. None of the treatment effect size estimates exceeds .25 standard deviation units, which demonstrates the treatment and control samples are balanced on observed covariates within the preferred bandwidths—providing evidence that assignment to treatment approximates random assignment in the region around the cutoff.

Table 1-8. Sample balance on standardized variables, conditional on forcing variable within optimal bandwidths

	2016		2017	
	β	SE	β	SE
Female	0.034	0.033	0.030	0.029
White	-0.192	0.014	-0.145	0.013
Black	0.224	0.030	0.191	0.026
Hispanic	-0.059	0.030	-0.131	0.027
Other race	0.106	0.008	0.103	0.007
Disabled	0.057	0.032	0.046	0.029
Gifted	0.184	0.027	0.127	0.025
Limited English proficiency	-0.101	0.033	-0.138	0.030
Over-age for grade	0.248	0.032	0.208	0.029
Nonstructural transfer in	0.021	0.034	0.030	0.030
Economically disadvantaged	-0.088	0.032	-0.077	0.028

NOTE: Estimates from RD with covariate listed in row as outcome and triangular kernel. Treatment and control samples within optimal CCT bandwidths.

Graphical analysis provides further evidence that the data meet the smoothness condition (Figure 1-2 and Figure 1-4), and we conduct an additional test in which we specify a series of placebo cutoffs and test for discontinuities. We find no evidence of significant discontinuities across multiple placebo cutoffs above and below the threshold in 2016 or 2017 (Table A-1-9).

Another assumption of the RD design is that student selection into or out of the treatment schools in response to the intervention is minimal. To test this assumption, we examine whether the demographics of NCT schools changed in response to the intervention in 2016 or 2017. Specifically, we estimate an RD on a set of school-level demographic characteristics in each of the two years of treatment within the optimal bandwidth. In the presence of student selection in and out of treatment schools, we would observe significant effects of NCT on these school-level demographics variables. Table A-1-10 shows there is no evidence for these selection effects.

As a final check, we test for differential attrition across the treatment and control schools. Three schools closed during the study period—one control and two treatment schools. Of those three schools, one treatment and one control school are within the optimal CCT bandwidth for both 2016 and 2017. The overall and differential levels of attrition both fall below the conservative boundary set in the What Works Clearinghouse standards (What Works Clearinghouse, 2017).

Table 1-9. Attrition

	CCT 2016 (4.13)	CCT 2017 (3.35)
β_{treat}	.042	.046
$\beta_{compare}$	.044	.048
$\beta_{overall}$	.043	.047
β_{diff}	-.002	-.003
(SE)	(.060)	(.066)

NOTE: Estimates from linear probability model predicting attrition at the school level and controlling for the forcing variable within the optimal CCT bandwidths and with a triangular kernel.

Discussion

We find that NCT reduced student achievement and increased teacher turnover and that the negative effects on student achievement may be associated with the timing of the CNA. While the increased teacher turnover in NCT schools in 2017 opens the possibility of strategic staffing by replacing less effective teachers with more effective ones, we find no evidence that strategic staffing occurred. Treatment schools experienced higher turnover across all levels of teacher effectiveness, with ineffective teachers no more likely to turn over than more effective teachers in their schools. Given NCT's largely rural context, this finding underscores the challenges associated with recruiting and retaining effective teachers in rural schools, which are unlikely to have robust educator labor markets from which to draw. Turnaround efforts that rely on strategic staffing may be less effective in rural contexts if they fail to counteract these labor market challenges with financial incentives that were a part of some effective turnaround efforts or other effective approaches aimed at recruiting and retaining effective teachers. Our findings that NCT schools did not recruit more effective teachers provide evidence against the possibility of a general equilibrium effect of targeted turnaround on nearby schools such as what occurred in Tennessee's iZones (Kho et al., 2019). However, unlike the iZones, which comprised urban schools, those dynamics may be expected to play out differently in rural schools that need to recruit teachers from outside the local area.

Under NCT, DPI provided coaching support for teachers and principals, but the amount of coaching varied across and within schools. Rather than building school capacity through strategic staffing and focusing on schoolwide processes and practices such as establishing a supportive and collaborative environment, NCT prioritized coaching to develop individual teacher skills and capacity in schools where, on average, the entire staff turns over every three years. Developing individual capacities may be an essential component of turnaround in rural schools, but our findings suggest it is not sufficient on its own—and on its face is unlikely to be an effective strategy unless complementary reforms are implemented to reduce the turnover of the teachers who have increased their instructional skills. Strategic staffing is less likely to be an effective strategy in this largely rural sample of schools than in urban or suburban schools that can draw from a larger pool of educators in the local labor market, especially without regulations and funds to support incentives for effective teachers to transfer into and remain in these low-performing schools. In fact, teachers in NCT schools expressed being stigmatized from working in a school labeled as “low performing.”

While we cannot know for certain whether the first two years of NCT laid the groundwork for incremental improvement in future years, we find no evidence that delayed positive effects are emerging. For example, the NCT theory of change focused largely on building the capacity of individual teachers and principals, but many of those teachers left NCT schools in 2017, taking any increased capacity with them. Additionally, because of the emphasis on individual-level capacities, it is unlikely that the intervention fostered the development of school-level systems and processes required to sustain long-term school improvement.

It is possible that targeting all schools in the bottom 5 percent produced negative effects by spreading resources too thin. Specifically, providing limited, inconsistent supports may have contributed to an already unstable school environment. Under ESSA, states are required to designate the bottom 5 percent of schools as low performing but are not necessarily required to serve the full 5 percent with the same reform model. Larger negative effects in the higher achieving of the lowest performing schools—beginning in the first year and increasing in the second year—suggest states might not be able or willing to allocate sufficient resources to effectively serve all schools in the lowest 5 percent of performance. In addition, the differential effects that appear to be associated with the conduct and timing of comprehensive needs assessments—which are mandated in the ESSA legislation—point to the importance of needs assessment timing and finding the resources, both human and financial, to conduct the needs assessments prior to the school improvement planning and implementation. Such efforts might require a planning year and additional human resources prior to initiating comprehensive services.

Three limitations are relevant to interpreting these findings. First, the regression discontinuity design focuses intentionally on schools around the eligibility cutoff in order to minimize threats to internal validity related to baseline differences between schools. While the findings are consistent across a wider set of bandwidths, the RD estimates represent the estimated effects of NCT for a narrow band of schools around the cutoff and the generalizability of the estimates is limited by the focus on these schools. Second, 21 of the schools receiving NCT services also received turnaround services under the state's RttT grant. Because these schools were in the bottom 5 percent in two different rounds of identification, it is possible that they may be more resistant to turnaround efforts and that the negative effects stem in part from that resistance. Finally, the generalizability of these findings should be considered in the context of the sample. The implementation of a theory of action that hindered student achievement in this sample of schools would not necessarily have the same effects in urban or suburban settings. However, low-performing schools are in rural, suburban, town, and urban contexts, and school turnaround under ESSA will target schools in each of these contexts. Additionally, many of the lessons learned under NCT are likely applicable beyond the rural context. For example, North Carolina made decisions to spread limited resources across a large number of schools and to rely on a theory of change that does not effectively transform school-level processes and practices nor promote strategic staffing practices. These strategies were included as part successful turnaround models in other states.

Conclusion

As states implement plans to support their lowest performing schools under ESSA, our findings suggest that school reform without intentional disruption of the status quo or supplemental resources has the potential to hinder student achievement and increase unintentional teacher turnover. This analysis also suggests that direct service provision without the backing of an influx of funding may not be a viable turnaround strategy across the entire set of schools in the bottom 5 percent in each state.

While these findings provide some descriptive evidence to explain the mechanisms underlying the negative effects of NCT on student achievement, future research could examine factors that may mediate or suppress the effects of interventions to improve the lowest performing schools. Such factors may include implementation fidelity and quality, school morale, and school climate.

References

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54–76. https://doi.org/10.3102/0162373716663646

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Atchison, D. (2020). The Impact of Priority School Designation Under ESEA Flexibility in New York State. *Journal of Research on Educational Effectiveness*, *13*(1), 121–146. https://doi.org/10.1080/19345747.2019.1679930

Bonilla, S., & Dee, T. (2017). *The Effects of School Reform Under NCLB Waivers: Evidence from Focus Schools in Kentucky* (Working Paper No. 23462). National Bureau of Economic Research. https://doi.org/10.3386/w23462

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *Stata Journal*, *17*(2), 372–404.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Carlson, D., & Lavertu, S. (2018). School Improvement Grants in Ohio: Effects on Student Achievement and School Administration. *Educational Evaluation and Policy Analysis*, 0162373718760218. https://doi.org/10.3102/0162373718760218

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

Dee, T., & Dizon-Ross, E. (2019). School Performance, Accountability, and Waiver Reforms: Evidence From Louisiana. *Educational Evaluation and Policy Analysis*, 0162373719849944. https://doi.org/10.3102/0162373719849944

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in Texas. *The Georgetown Public Policy Review*, 21–39.

Dougherty, S. M., & Weiner, J. M. (2017). The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance. *Educational Policy*, 0895904817719520. https://doi.org/10.1177/0895904817719520

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., & Giffin, J. (2017). *School Improvement Grants: Implementation and Effectiveness. NCEE 2017-4013*. National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED572215

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. https://doi.org/10.3102/0162373709333886

Hanushek, E. A., Rivkin, S. G., & Schiman, J. C. (2016). Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review*, *55*, 132–148. https://doi.org/10.1016/j.econedurev.2016.08.004

Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, *62*, 302–320. https://doi.org/10.1016/j.econedurev.2017.08.001

Hemelt, S. W., & Jacob, B. (2017). *Differentiated Accountability and Education Production: Evidence from NCLB Waivers* (Working Paper No. 23461). National Bureau of Economic Research. https://doi.org/10.3386/w23461

Hemelt, S. W., & Jacob, B. A. (2018). How Does an Accountability Program that Targets Achievement Gaps Affect Student Performance? *Education Finance and Policy*, 1–68. https://doi.org/10.1162/edfp_a_00276

Henry, G. T., Campbell, S. L., Thompson, C. L., & Townsend, L. W. (2014). *Evaluation of District and School Transformation School-Level Coaching and Professional Development Activities*.

Henry, G. T., Guthrie, J. E., & Townsend, L. W. (2015). *Outcomes and Impacts of North Carolina's Initiative to Turn Around the Lowest-Achieving Schools*. http://cerenc.org/wp-content/uploads/2015/09/ES-FINAL-Final-DST-Report-9-3-15.pdf

Henry, G. T., Pham, L. D., Kho, A., & Zimmer, R. (2020). Peeking Into the Black Box of School Turnaround: A Formal Test of Mediators and Suppressors. *Educational Evaluation and Policy Analysis*, *42*(2), 232–256. https://doi.org/10.3102/0162373720908600

Henry, G. T., & Redding, C. (2018). The consequences of leaving school early: The effects of within-year and end-of-year teacher turnover. *Education Finance and Policy*, 1–52.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., & Redding, S. (2008). *Turning Around Chronically Low-Performing Schools: A Practice Guide* (NCEE 2008-4020; IES Practice Guide). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/PracticeGuide.aspx?sid=7

Imbens, G., & Kalyanaraman, K. (2009). *Optimal Bandwidth Choice for the Regression Discontinuity Estimator* (Working Paper No. 14726). National Bureau of Economic Research. https://doi.org/10.3386/w14726

Imbens, G., & Lemieux, T. (2007). *Regression Discontinuity Designs: A Guide to Practice* (Working Paper No. 13039). National Bureau of Economic Research. https://doi.org/10.3386/w13039

Kho, A., Henry, G. T., Zimmer, R., & Pham, L. (2019). *General Equilibrium Effects of Recruiting High-Performing Teachers for School Turnaround: Evidence from Tennessee* (EdWorkingPaper: 19-64). Annenberg Institute at Brown University. https://edworkingpapers.com/ai19-64

Papay, J., & Hannon, M. (2018, November 8). *The Effects of School Turnaround Strategies in Massachusetts*. 2018 APPAM Fall Research Conference: *Evidence for Action: Encouraging Innovation and Improvement*, Washington, D.C. https://appam.confex.com/appam/2018/webprogram/Paper26237.html

Peurach, D., & Neumerski, C. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379–420. https://doi.org/10.1007/s10833-015-9259-z

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, *50*(1), 4–36. https://doi.org/10.3102/0002831212463813

Schueler, B. E., Goodman, J., & Deming, D. J. (2016). *Can States Take Over and Turn Around School Districts? Evidence from Lawrence, Massachusetts* (Working Paper No. 21895). National Bureau of Economic Research. http://www.nber.org/papers/w21895

Stock, J. H., & Yogo, M. (2002). *Testing for Weak Instruments in Linear IV Regression* (Working Paper No. 284). National Bureau of Economic Research. https://doi.org/10.3386/t0284

Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, S., & Weinstein, T. (2016). The Impact of Turnaround Reform on Student Outcomes: Evidence and Insights from the Los Angeles Unified School District. *Education Finance and Policy*, *11*(3), 251–282. https://doi.org/10.1162/EDFP_a_00188

Sun, M., Penner, E. K., & Loeb, S. (2017). Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants. *American Educational Research Journal*, *54*(4), 607–643. https://doi.org/10.3102/0002831217695790

What Works Clearinghouse. (2017). *Standards Handbook. Version 4.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Zimmer, R., Henry, G. T., & Kho, A. (2017). The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones. *Educational Evaluation and Policy Analysis*, *39*(4), 670–696. https://doi.org/10.3102/0162373717705729

Appendix

Table A-1-1. Examination of treatment receipt by compliance

Panel A. Comparison of no-shows and control group compliers

	School transformation coaching				Instructional coaching			
	No coaching		Got coaching		No coaching		Got coaching	
	Sum	Percent	Sum	Percent	Sum	Percent	Sum	Percent
No-shows	3	60.0	2	40.0	2	40.0	3	60.0
Control group compliers	33	63.5	19	36.5	24	51.1	23	48.9
Chi-sq	0.023				0.221			
p-value	0.878				0.638			

Panel B. Comparison of always-takers and treatment group compliers

	School transformation coaching				Instructional coaching			
	No coaching		Got coaching		No coaching		Got coaching	
	Sum	Percent	Sum	Percent	Sum	Percent	Sum	Percent
Always-takers	0	0.0	6	100.0	0	0.0	6	100.0
Treatment group compliers	4	6.9	54	93.1	5	8.9	51	91.1
Chi-sq	0.441				0.583			
p-value	0.506				0.445			

Calculations from principal survey data. Survey question for column with school transformation coaching was, “Since January 2016, did you meet in-person, one-on-one with a school transformation coach or someone who has provided you with deliberate, sustained assistance designed to help you learn or figure out how to improve your current school?” Response options were (1) Yes, I received School Transformation Coaching from NC DPI, (2) Yes, I received advice/guidance/coaching from a source other than NC DPI, and (3) No. Responses of (1) and (2) were both coded as having received coaching, while a response of (3) was coded as not having received coaching. The response rate for this question was 81% for principals of schools assigned to treatment and 70% for principals of schools not assigned to treatment. Survey question for column with instructional coaching was “Have any of your teachers received in-person instructional coaching within your school building since January 2016?” Response options were (1) Yes, (2) No, and (3) I don’t know. A response of (1) was coded as having received coaching, a response of (2) was coded as not having received coaching, and a response of (3) was coded as missing. The response rate for this question 78% for principals of schools assigned to treatment and 64% for principals of schools not assigned to treatment.

Table A-1-2. Tests for validity of implementation groupings

Panel A: Comprehensive Needs Assessments and CNA unpackings

	(1) CNA timing	(2) Unpacking timing	(3) Unpacking presence
2014 or 2015	-3.667		
	(2.137)		
Spring 2016	-0.447		
	(1.976)		
2016-17 school year	2.221		
	(2.085)		
2014 or 2015		-4.360^{*}	
		(1.922)	
Summer 2016		-0.541	
		(1.708)	
2016-2017 during SY		-0.0667	
		(2.306)	
Unpacking occurred			-1.307
			(1.496)
Constant	27.38^{***}	28.00^{***}	27.82^{***}
	(1.582)	(1.122)	(1.209)
R^2	0.112	0.0752	0.0104
Obs	75	75	75

Estimates from regressions of 2015 performance composite on CNA timing group, unpacking presence group, and unpacking timing group, respectively. Reference categories are no CNA/pre-2014 CNA group, no unpacking group, and no unpacking/pre-2014 unpacking group, respectively. Standard errors in parentheses.

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

Panel B: Coaching dosage

	(1) Total coaching	(2) Instructional coaching	(3) School transformation coaching
Bottom quartile	3.416	-2.017	0.675
	(1.743)	(1.729)	(1.766)
Highest quartile	-1.509	-4.044*	-2.373
	(1.650)	(1.644)	(1.680)
Constant	26.59***	28.58***	27.46***
	(0.969)	(0.998)	(1.020)
R^2	0.0834	0.0788	0.0384
Obs	75	75	75

Estimates from regressions of 2015 performance composite on coaching dosage group. Reference category is middle 50% of schools. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-1-3. TOT estimates within alternative bandwidths and full sample *(outcome=test score)*

Panel A: 2016

	(1) No BW	(2)	(3) IK	(4)	(5) 200% IK	(6)
TOT	-0.027	-0.017	-0.186***	1.095	-0.146**	-0.086*
	(0.0478)	(0.0437)	(0.0526)	(0.5948)	(0.0527)	(0.0381)
Covariates		X		X		X
Bandwidth	23.0	23.0	0.9	0.9	1.7	1.7
First-stage *F*-stat	160.28	156.25	76.91	3.69	31.58	30.69
N	83896	83896	195437	195437	195437	195437
N Bandwidth	83896	83896	10184	10184	20909	20909
T schools in BW	78	78	12	12	20	20
C schools in BW	80	80	5	5	15	15

Panel B: 2017

	(1) No BW	(2)	(3) IK	(4)	(5) 200% IK	(6)
TOT	-0.110**	-0.088*	-0.307***	0.042*	-0.207***	-0.413***
	(0.0417)	(0.0427)	(0.0823)	(0.0173)	(0.0606)	(0.0716)
Covariates		X		X		X
Bandwidth	23.0	23.0	0.7	0.7	1.5	1.5
First-stage *F*-stat	158.76	144.24	48.86	5264.95	29.59	62.88
N	83393	83393	195099	195099	195099	195099
N Bandwidth	83393	83393	8473	8473	16740	16740
T schools in BW	78	78	11	11	15	15
C schools in BW	79	79	4	4	12	12

NOTE: Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All models include lagged score and subject fixed effects on the right side, with math as the reference category. 50% IK not included because the bandwidth size—which unlike the CCT procedure does not account for the clustering of students within schools—includes only three schools above the cutoff. Red outlines denote first-stage *F* statistics on the treatment indicator smaller than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage.

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

Table A-1-4. TOT estimates by subject *(outcome=test score)*

Panel A: Math

	2016			2017		
	(1) CCT	(2) 50% CCT	(3) 200% CCT	(4) CCT	(5) 50% CCT	(6) 200% CCT
TOT	-0.098	-0.141**	-0.053	-0.096	-0.159*	-0.117*
	(0.0509)	(0.0513)	(0.0489)	(0.0698)	(0.0750)	(0.0576)
Covariates	no	no	no	no	no	no
Bandwidth	4.1	2.1	8.3	3.3	1.7	6.7
First-stage *F*-stat	79.92	34.57	199.37	79.57	29.48	212.87
N	85131	85131	85131	85130	85130	85130
N Bandwidth	21766	10039	39688	17026	8086	33235
T schools in BW	36	22	66	31	18	55
C schools in BW	51	19	102	37	13	84

Panel B: Reading

	2016			2017		
	(1) CCT	(2) 50% CCT	(3) 200% CCT	(4) CCT	(5) 50% CCT	(6) 200% CCT
TOT	-0.031	-0.094	0.002	-0.164***	-0.242***	-0.129***
	(0.0567)	(0.0614)	(0.0418)	(0.0370)	(0.0517)	(0.0327)
Covariates	no	no	no	no	no	no
Bandwidth	4.1	2.1	8.3	3.3	1.7	6.7
First-stage *F*-stat	79.92	22.94	153.76	54.91	19.10	154.75
N	88535	88535	88535	88421	88421	88421
N Bandwidth	22436	10420	41286	17611	8312	34617
T schools in BW	36	22	66	31	18	55
C schools in BW	51	19	102	37	13	84

Panel C: Science

	2016			2017		
	(1)	(2)	(3)	(4)	(5)	(6)
	CCT	50% CCT	200% CCT	CCT	50% CCT	200% CCT
TOT	-0.072	-0.326**	-0.043	-0.142	-0.187	-0.045
	(0.1658)	(0.1219)	(0.1075)	(0.1290)	(0.1491)	(0.1056)
Covariates	no	no	no	no	no	no
Bandwidth	4.1	2.1	8.3	3.3	1.7	6.7
First-stage *F*-stat	901.80	51810.86	924.16	1109.56	6.9224e+29	1455.42
N	21771	21771	21771	21548	21548	21548
N Bandwidth	6529	2956	11540	4786	2226	9568
T schools in BW	33	20	56	28	17	48
C schools in BW	50	18	97	37	13	81

NOTE: Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All models include lagged score on the right side. Red outlines denote first-stage F statistics on the treatment indicator smaller than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage.

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

Table A-1-5. TOT estimates without lagged test score *(outcome=test score levels)*

	2016			2017		
	(1)	(2)	(3)	(4)	(5)	(6)
	CCT	50% CCT	200% CCT	CCT	50% CCT	200% CCT
TOT	-0.054	-0.209	-0.019	-0.210	-0.429***	-0.142
	(0.1041)	(0.1120)	(0.0686)	(0.1260)	(0.1262)	(0.0872)
Covariates	no	no	no	no	no	no
Bandwidth	4.1	2.1	8.3	3.3	1.7	6.7
First-stage *F*-stat	72.76	20.16	122.99	47.89	16.00	132.48
N	235611	235611	235611	234659	234659	234659
N Bandwidth	59238	27245	109730	45948	21580	91816
T schools in BW	36	22	66	31	18	55
C schools in BW	51	19	102	37	13	84

NOTE: Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All models include subject fixed effects on the right side, with math as the reference category.

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

Table A-1-6. TOT estimates by school level *(outcome=test score)*

Panel A: Elementary

	2016			2017		
	(1) CCT	(2) 50% CCT	(3) 200% CCT	(4) CCT	(5) 50% CCT	(6) 200% CCT
TOT	-0.025	0.039	-0.033	-0.326	-0.640	-0.293**
	(0.0785)	(0.2938)	(0.0591)	(0.1978)	(0.8842)	(0.1061)
Covariates	no	no	no	no	no	no
Bandwidth	4.1	2.1	8.3	3.3	1.7	6.7
First-stage *F*-stat	3.13	0.50	6.76	2.62	0.38	7.18
N	54933	54933	54933	56572	56572	56572
N Bandwidth	10510	4896	22309	8623	4124	20234
T schools in BW	20	10	34	16	7	29
C schools in BW	20	9	50	15	7	41

Panel B: Middle

	2016			2017		
	(1) CCT	(2) 50% CCT	(3) 200% CCT	(4) CCT	(5) 50% CCT	(6) 200% CCT
TOT	-0.069	-0.143**	-0.033	-0.090	-0.123***	-0.078
	(0.0610)	(0.0442)	(0.0546)	(0.0493)	(0.0311)	(0.0457)
Covariates	no	no	no	no	no	no
Bandwidth	4.1	2.1	8.3	3.3	1.7	6.7
First-stage *F*-stat	303.46	2.43049e+35	381.81	442.26	5.184e+35	598.29
N	124063	124063	124063	122863	122863	122863
N Bandwidth	34957	16508	57805	27513	13488	48029
T schools in BW	12	9	20	12	9	16
C schools in BW	24	8	36	17	5	31

Panel C: High [a]

	2016		2017	
	(1) CCT	(2) 200% CCT	(3) CCT	(4) 200% CCT
TOT	0.022	-0.001	-0.199***	-0.112*
	(0.0428)	(0.0343)	(0.0362)	(0.0561)
Covariates				
Bandwidth	4.1	8.3	3.3	6.7
N	16441	16441	15664	15664
N Bandwidth	5264	12400	3287	9157
T schools in BW	4	12	3	10
C schools in BW	7	16	5	12

NOTE: Elementary and middle schools are estimated using fuzzy RD. High school models use a sharp RD because there is no noncompliance at the high school level.

Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All models include lagged score and subject on the right side, with math as the reference category. Red outlines denote first-stage F statistics on the treatment indicator smaller than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] High schools only estimated within CCT bandwidth and 200% CCT bandwidth because there are not enough high schools within the 50% bandwidth.

Table A-1-7. TOT estimates within alternative bandwidths and full sample ***(outcome=teacher turnover)***

Panel A: 2016

	(1)	(2)	(5)	(6)	(7)	(8)
	No BW		IK		200% IK	
TOT	-0.075	-0.093	0.332*	0.078	0.152	0.331
	(0.0586)	(0.0544)	(0.1377)	(0.0792)	(0.1243)	(0.2055)
Covariates		X		X		X
Bandwidth	23.0	23.0	0.9	0.9	1.7	1.7
First-stage *F*-stat	78.50	76.91	12.04	11.02	6.10	3.84
N	4783	4783	10770	10770	10770	10770
N Bandwidth	4783	4783	488	488	1032	1032
T schools in BW	76	76	12	12	19	19
C schools in BW	80	80	5	5	15	15

Panel B: 2017

	(1)	(2)	(5)	(6)	(7)	(8)
	No BW		IK		200% IK	
TOT	0.099	0.120*	0.179*	0.056	0.378**	0.470
	(0.0511)	(0.0478)	(0.0706)	(0.0527)	(0.1453)	(0.3481)
Covariates		X		X		X
Bandwidth	23.0	23.0	0.7	0.7	1.5	1.5
First-stage *F*-stat	80.46	78.68	18.06	390.46	6.25	4.41
N	4707	4707	10492	10492	10492	10492
N Bandwidth	4707	4707	424	424	844	844
T schools in BW	76	76	11	11	15	15
C schools in BW	79	79	4	4	12	12

NOTE: Estimates from linear probability models. Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. 50% IK not included because the bandwidth size—which unlike the CCT procedure does not account for the clustering of students within schools—includes only three schools above the cutoff. Red outlines denote first-stage *F*-statistics on the treatment indicator smaller than the What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage. IK bandwidths calculated using the fuzzy test score models.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-1-8. ITT estimates *(outcome=teacher turnover)*

Panel A: 2016

	(1)	(2)	(3)	(4)	(5)	(6)
	CCT		50% CCT		200% CCT	
ITT	-0.036	-0.075	0.066	0.068	-0.059	-0.067
	(0.0777)	(0.0610)	(0.1016)	(0.0487)	(0.0465)	(0.0411)
Covariates		X		X		X
Bandwidth	4.1	4.1	2.1	2.1	8.3	8.3
N	10770	10770	10770	10770	10770	10770
N Bandwidth	2658	2658	1240	1240	5270	5270
T schools in BW	35	35	21	21	64	64
C schools in BW	51	51	19	19	102	102

Panel B: 2017

	(1)	(2)	(3)	(4)	(5)	(6)
	CCT		50% CCT		200% CCT	
ITT	0.187^{**}	0.138^{**}	0.258^{***}	0.143^{**}	0.104	0.096^{*}
	(0.0668)	(0.0507)	(0.0620)	(0.0508)	(0.0554)	(0.0460)
Covariates		X		X		X
Bandwidth	3.3	3.3	1.7	1.7	6.7	6.7
N	10492	10492	10492	10492	10492	10492
N Bandwidth	2078	2078	940	940	4280	4280
T schools in BW	30	30	17	17	53	53
C schools in BW	37	37	13	13	84	84

NOTE: Estimates from linear probability models. Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-1-9. Placebo estimates from fuzzy RD within optimal CCT bandwidth, 2016 *(outcome=test score)*

Panel A: 2016

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Placebo Cutoff	-4	-3	-2	-1	1	2	3	4
TOT	-2.778	-0.200	0.058	0.058	0.171	-0.018	0.187	0.782
	(51.5073)	(1.3754)	(0.1058)	(0.1929)	(0.1863)	(0.1664)	(0.1067)	(7.6192)
Observations	195466	195466	195466	195466	195466	195466	195466	195466

Panel B: 2017

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	-4	-3	-2	-1	1	2	3	4
TOT	-0.173	1.085	-0.040	-0.136	0.019	0.004	-0.103	5.101
	(1.8885)	(14.6048)	(0.1941)	(0.1569)	(0.2192)	(0.1951)	(0.1516)	(65.7928)
Observations	195078	195078	195078	195078	195078	195078	195078	195078

NOTE: Standard errors clustered at the school level. All models include lagged score and subject on the right side, with math as the reference category. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-1-10. School demographics by treatment year

	2016			2017		
	Treat	Control	p-value	Treat	Control	p-value
ED percent	65.34	68.62	0.561	68.32	65.37	0.780
Minority percent	77.60	74.34	0.702	78.20	67.41	0.340
Black percent	48.72	48.89	0.988	49.45	42.41	0.650
Hispanic percent	16.72	20.11	0.583	17.73	20.33	0.710
ADM	418.23	433.38	0.885	399.75	428.53	0.808

NOTE: Estimates from RD predicting covariate listed in row as outcome and triangular kernel. Treatment and control samples within optimal CCT bandwidths using a triangular kernel.

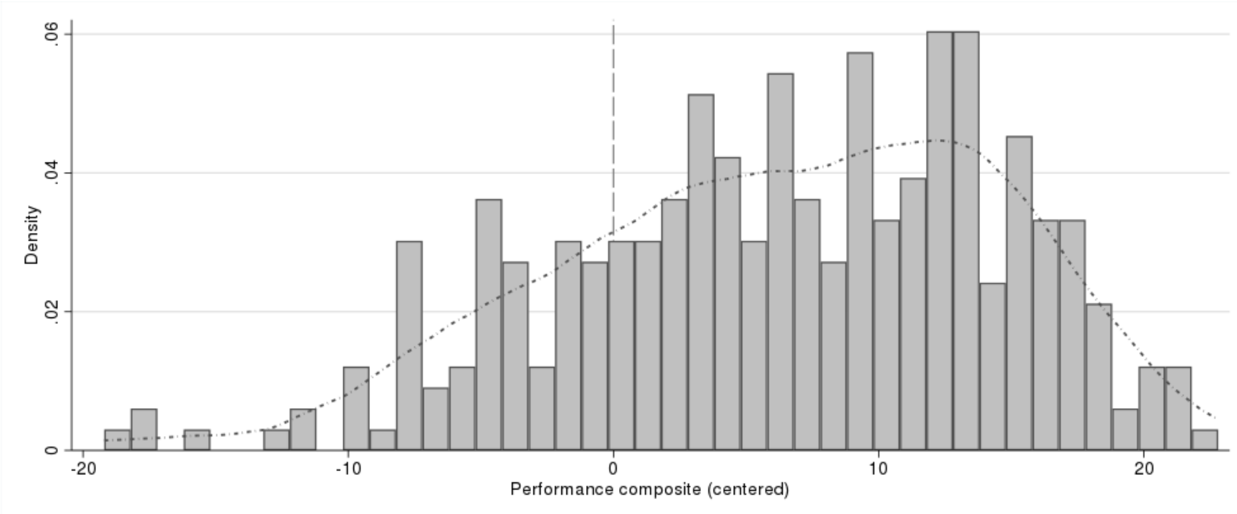
Table A-1-11. Fuzzy RD results by CNA timing *(outcome=test score)*

	2016		2017	
	(1)	(2)	(3)	(4)
	Full sample	CCT	Full sample	CCT
Pre-2014 or none	-0.042	-0.100	-0.203^{**}	-0.225^{***}
	(0.0576)	(0.0903)	(0.0754)	(0.0674)
2014 or 2015	-0.091^{**}	-0.104	-0.087^{**}	-0.168^{***}
	(0.0324)	(0.0653)	(0.0333)	(0.0458)
Spring 2016	-0.027	-0.028	-0.114^{***}	-0.144^{*}
	(0.0360)	(0.0710)	(0.0330)	(0.0630)
2016-17 school year	-0.004	-0.040	-0.029	-0.067
	(0.0302)	(0.0647)	(0.0252)	(0.0471)
Constant	-0.103^{***}	-0.078	-0.102^{***}	-0.075^{*}
	(0.0188)	(0.0450)	(0.0181)	(0.0354)
N	86354	51969	85808	39427

NOTE: 2SLS estimates from fuzzy RD using triangular kernel with four separate treatments by CNA timing. Standard errors clustered at the school level. CCT bandwidths calculated using the fuzzy test score models. All first-stage F-statistics are greater than What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage. All models include lagged score and subject fixed effects on the right side, with math as the reference category.

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

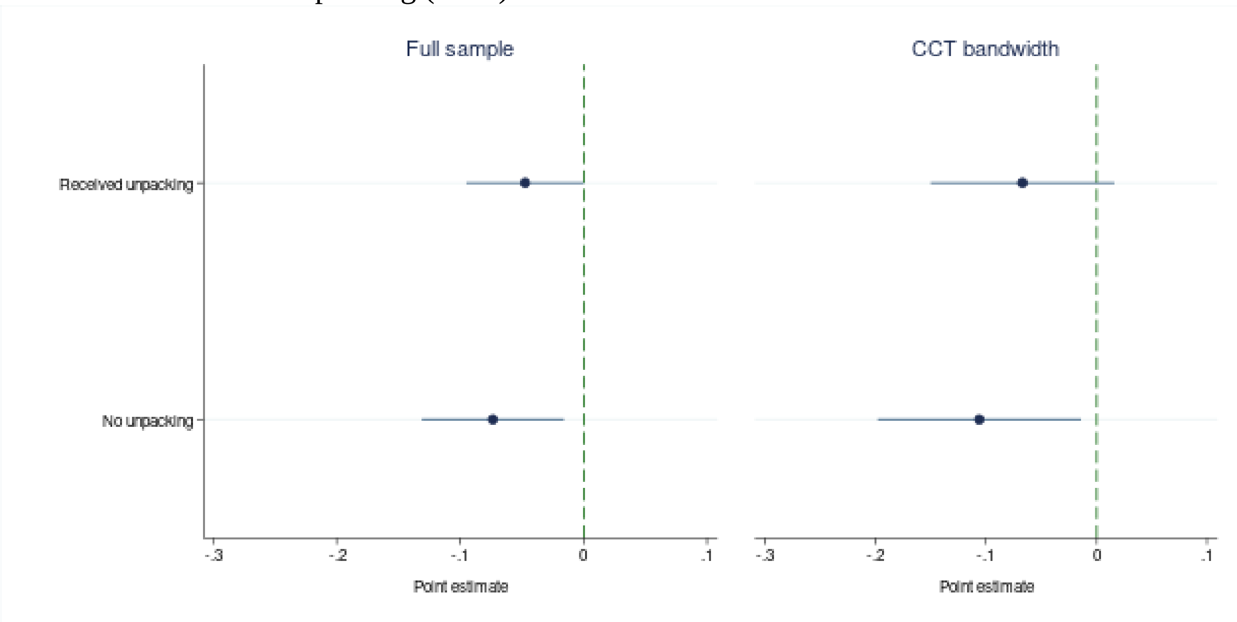
Figure A-1-1. Graphical integrity of the forcing variable



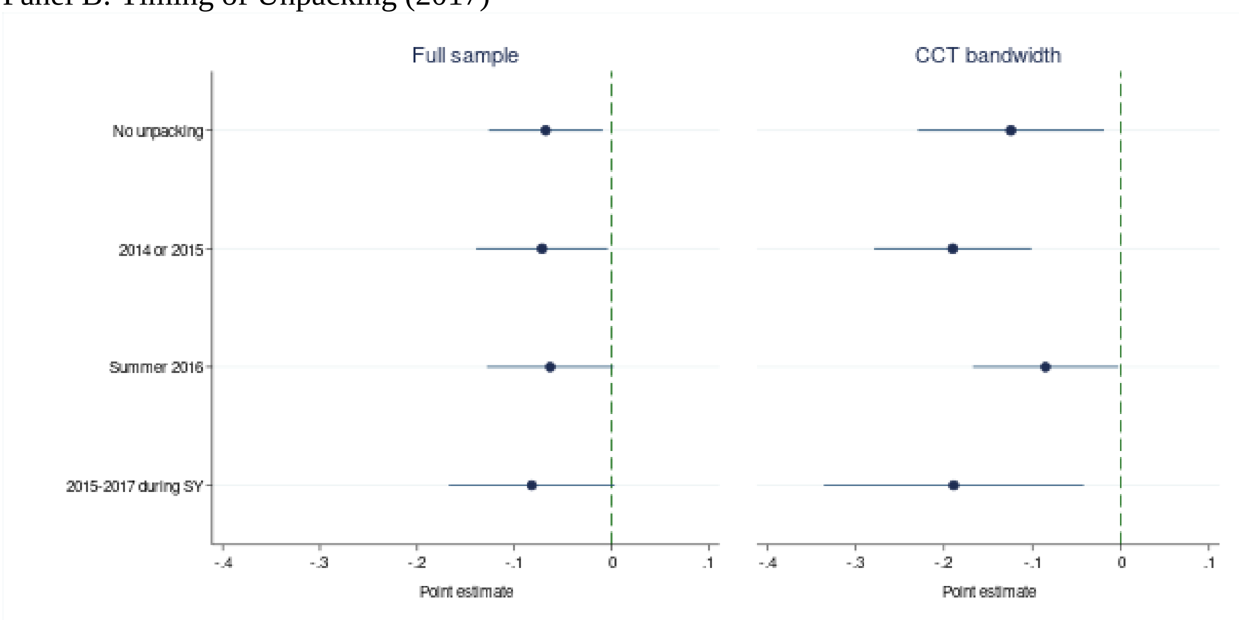
NOTE: Bin width is 1. Includes all eligible schools.

Figure A-1-2. Estimated effects by presence of CNA unpacking

Panel A: Presence of Unpacking (2017)



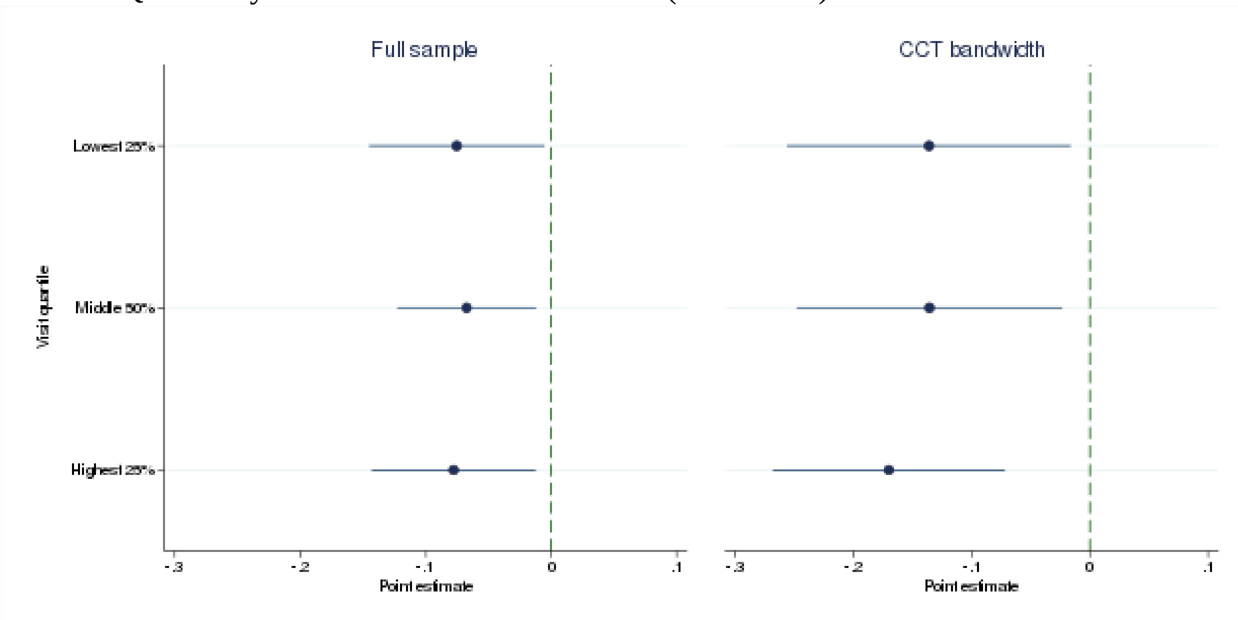
Panel B: Timing of Unpacking (2017)



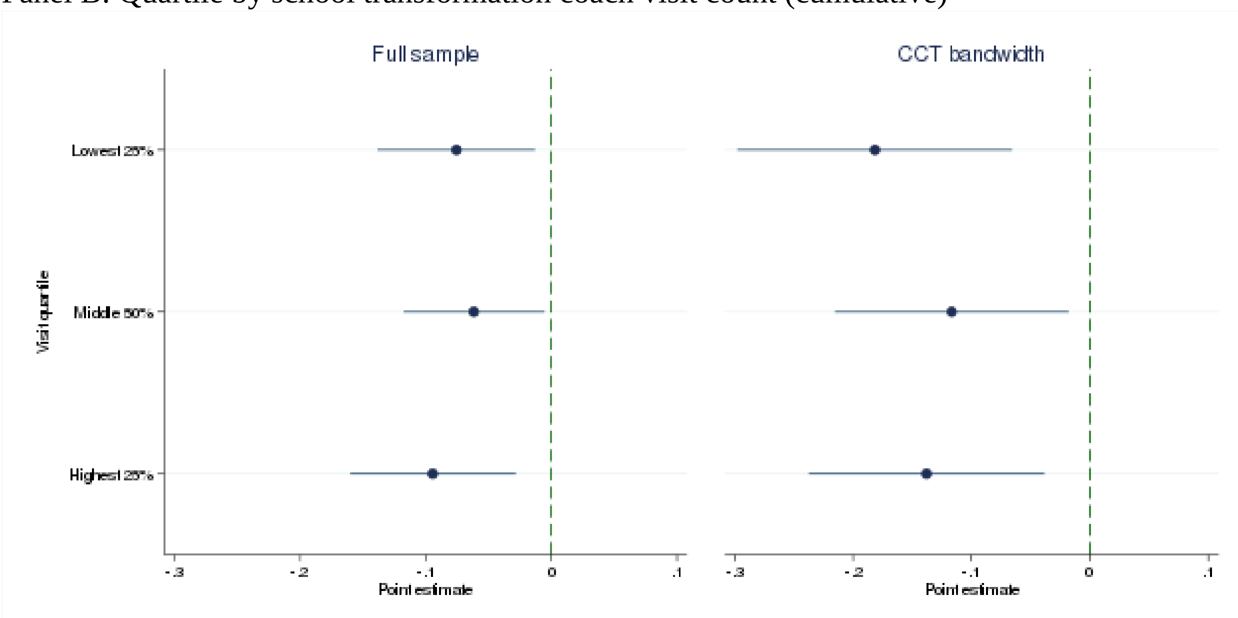
NOTE: 2SLS estimates from fuzzy RD using triangular kernel with separate treatments by CNA unpacking presence (panel A) and timing (panel B). All models include lagged score and subject on the right side, with math as the reference category. All first-stage *F*-statistics are greater than What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage. Preferred CCT bandwidths from fuzzy test score models. Standard errors clustered at the school level.

Figure A-1-3. Estimated effects by coaching dosage

Panel A: Quartile by instructional coach visit count (cumulative)



Panel B: Quartile by school transformation coach visit count (cumulative)



NOTE: 2SLS estimates from fuzzy RD using triangular kernel with separate treatments for schools in the bottom quartile of number of visits, middle 50% of number of visits, and top quartile of number visits. Quartiles by school level. All first-stage *F*-statistics are greater than What Works Clearinghouse (2017) recommended minimum size of 16 for a sufficiently strong first stage. Preferred CCT bandwidths from fuzzy test score models. Standard errors clustered at the school level.

Chapter 2

An Early Warning System for Low-Performing Schools: Developing a Multidimensional Measure of Risk

In the years since the Elementary and Secondary Education Act (ESEA) began requiring states to identify and support their lowest performing schools, new schools have fallen into the lowest performing group each year. While some schools languish in the bottom 5 percent year after year, others cycle in and out of the lowest performing category or even drop in for a single year. The most recent reauthorization of ESEA, the Every Student Succeeds Act (ESSA, 2015), includes a mandate that states continue to identify and attempt to reform their lowest performing schools, ensuring that every state will continue to grapple with turning around low performing schools into the foreseeable future. States are required to provide supports to these low-performing schools, classified as Comprehensive Support and Improvement (CSI) schools under ESSA. ESSA expanded the definition of low performing beyond proficiency rates by requiring states to use at least one non-academic indicator in calculating a school performance score, opening a window for state education leaders to rethink what constitutes school underperformance and the mechanisms through which schools fail. This shift in thinking is noteworthy because indices derived from student test score data require schools to fail before they can be targeted for supports.

A small body of literature has attempted to identify what characterizes school decline, defining decline in terms of student achievement over time (Hochbein, 2011, 2012b, 2012a; Hochbein & Duke, 2011). This narrow definition of decline mirrors federal accountability policy under No Child Left Behind (NCLB, 2001), which shed light on low-performing schools through increased testing and required reporting toward performance goals. But declining test scores are a symptom of failure—not the cause or even a leading indicator. Leading indicators of failure occur before test scores begin to fall. Existing research suggests that three dimensions of schooling are associated with a school's ability to be resilient in the presence of school-level shocks that may undermine student learning. These dimensions, each of which are supported by prior literature, are student preparedness and supports, school climate, and teachers and leaders (Darling-Hammond, 2000; Grissom, 2011; Hochbein & Duke, 2011; Kraft & Papay, 2014; Leithwood, Harris, & Strauss, 2010; Ronfeldt, Farmer, McQueen, & Grissom, 2015; Ronfeldt, Loeb, & Wyckoff, 2013).

This dissertation essay seeks to understand the underlying causes of school underperformance so schools can be identified and supported before they fail. The goal of this dissertation essay is threefold. First, I undertake a descriptive analysis mapping the trajectories of schools into the bottom 5 percent and describe how those trajectories as well as the stability of schools designated as low performing vary by performance index. Second, I identify and examine the correlates of school underperformance following the framework of the three dimensions above. The third goal was to develop a measure of a school's risk of becoming low performing that predicts low performance before failure occurs, though as I describe below the correlates of school underperformance do not load well onto a composite risk measure.

Literature Review

School turnaround and improvement

Existing research has shown there is substantial heterogeneity in the effects of school turnaround efforts. Empirical evaluations of the first federally funded school improvement program, the Comprehensive School Reform Demonstration program (CSRD), found CSRD programs did not produce positive average effects on student achievement and in fact may have led to lower math performance among black and Hispanic students (Gross, Booker, & Goldhaber, 2009). In 2008, School Improvement Grants (SIG) and Race to the Top (RttT) introduced school turnaround to the school improvement toolkit, with the distinction that turnaround would create dramatic and rapid change in chronically low-performing schools whereas the previous generation of school reform models assumed an incremental approach (Peurach & Neumerski, 2015). Whole-school reform continued with No Child Left Behind (NCLB waivers) in which states were required to identify "priority" and "focus" schools for intervention but did not need to follow any of the four models that were prescribed under SIG and RttT, marking a shift to a flexible school reform mandate that would continue under the Every Student Succeeds Act (ESSA) without the dedicated funding to undertake those turnaround efforts.

School reform programs under SIG, RttT, and NCLB waivers have yielded mixed effects. A large-scale evaluation of SIG grants in 22 states found schools implementing SIG models did not increase test scores, high school graduation, or college enrollment (Dragoset et al., 2017), and a study of SIG schools in Texas found negative effects on student achievement in elementary and middle schools and no effect on student achievement in high schools (Dickey-Griffith, 2013). While these unsuccessful efforts have varied in approach from state to state and even within states, some researchers have hypothesized that a common thread across interventions is a failure to address the larger educational infrastructure—the basic foundational structures, systems, and resources (Peurach & Neumerski, 2015)—in which school failure has occurred (Dougherty & Weiner, 2017). School improvement efforts that neglect the underlying infrastructure problems may be unsuccessful because they fail to address the school's barriers to improvement, such as high educator and student mobility, student absenteeism, inadequate processes for collaboration, and others (Henry & Harbatkin, 2019; Henry, Pham, Kho, & Zimmer, 2020). Even reforms that might yield some early successes are unlikely to sustain progress in the longer term if they do not address deficiencies in schools' educational infrastructure (Cohen & Moffitt, 2010; Peurach & Neumerski, 2015).

Across four studies using regression discontinuity (RD) to examine the effects of being designated a Focus School under NCLB waiver reforms in four states, only one showed positive effects on student achievement. In Kentucky, Focus School reforms led to significant increases math and reading proficiency among so-called "gap group," students—which are those students who qualify for free or reduced price lunch, special education, or are black, Hispanic, American Indian, or Limited English proficient (LEP)—and potentially higher proficiency among non-gap group students. The authors suggested that Kentucky's focus on these gap group students combined with a clearly articulated set of reform activities from the state contributed to uniquely positive effects of the state's waiver-driven reforms (Bonilla & Dee, 2017).

An evaluation of Focus Schools in Michigan, where the state assigned schools to Focus status based on the achievement gap between the top and bottom 30 percent of students, found no effect on average math and reading scores in each of four years following the Focus designation (Hemelt & Jacob, 2018). The intervention, which was funded through a combination of existing building-level and district-specific Title 1 funds, involved employing locally tailored strategies based on district-led conversations about school data. The intervention led to a significant decrease in the math achievement gap in the first year, though that reduction appeared to be driven by test score declines among the higher achieving students rather than improvements among the lower achieving students. Meanwhile, an evaluation of Michigan's Priority Schools—the state's bottom 5 percent based on a composite score composed of achievement levels, achievement growth, and the achievement gap between the highest and lowest students—found no consistent effects on math and reading scores and some negative effects on average reading scores one and two years after the Priority designation (Hemelt & Jacob, 2017). While Priority Schools in Michigan were required to follow one of the four federal turnaround models, they did not receive the influx of ARRA funds that RttT and SIG schools received and instead were required to set aside 10 percent of building Title 1 funds toward the reform.

Focus School reforms in Louisiana had no consistent effect on school performance over three years, though there is suggestive evidence that school performance declined in the third year of supports. In contrast to Michigan and Kentucky, Louisiana identified Focus Schools using the overall school performance score rather than subgroup performance or achievement gaps. The Louisiana reforms included the adoption of well-publicized school letter grades based on school performance and called for tailored interventions based on school needs as determined by a needs assessment (Dee & Dizon-Ross, 2019). This second component of the Focus School intervention aligns with requirements under ESSA, potentially highlighting the importance of correctly identifying the determinants of low performance and accurately diagnosing those problems in a low-performing school.

In Rhode Island, a school turnaround model that provided districts with autonomy to select from a menu of interventions had no immediate effect on student achievement, and schools that were required to implement more of these interventions experienced declines in reading test scores in the second and third years of the reform (Dougherty & Weiner, 2017). Meanwhile, there was no effect of being designated a "Warning School," the next tier of schools targeted for intervention. Schools in Rhode Island were identified using a composite score based on proficiency, growth, and subgroup performance, in addition to graduation rates for high schools and testing participation rates.

But emerging research is beginning to identify effective turnaround interventions in Massachusetts, California, and Tennessee under RttT and SIG (e.g., Dee, 2012; Papay, 2015; Zimmer, Henry, & Kho, 2017). In Massachusetts, which focused turnaround efforts on its 35 lowest performing schools, RD and comparative interrupted time series analyses found significant positive effects in math and ELA achievement for students in these schools. These effects emerged in the first year of turnaround and grew through the fourth year. The authors found suggestive evidence that the large positive effects stemmed in part from deselection of low-performing teachers and improvements of teachers who stayed in these schools—suggesting the possibility for a turnaround intervention to improve student achievement over multiple years by investing in human capital within the school (Papay & Hannon, 2018). Also in Massachusetts, a state takeover of Lawrence Public Schools in 2011 yielded positive results in math and reading for students during the first two years of the intervention (Schueler, Goodman, & Deming, 2017). While the state takeover of Lawrence was not funded by SIG, some of the practices it employed aligned with SIG models, including replacing 56 percent of district principals over two years and 10 percent of teachers in the first year, setting ambitious annual accountability targets, and expanding learning time (Schueler et al., 2017).

Policymakers in Tennessee implemented two distinct school reform models under RttT: the Achievement School District (ASD) removed low-performing schools from their local districts and placed them under the auspices of a charter management organization or the state, while the Innovation Zones (iZones) kept schools in their local education agencies but created semiautonomous districts-within-districts that supported schools by helping them to attract, retain, and develop high quality teachers and leaders. Research on school turnaround in Tennessee found significant increases in math, reading, and science achievement in iZones but no significant effects for students in ASD schools (Zimmer et al., 2017). The authors found that iZone schools, which offered substantial raises to its teachers, were more effective than ASD schools at developing and retaining high value-added teachers. Similar to the Massachusetts effort, the iZone intervention provides evidence that strategic investments designed to address barriers associated with attracting and retaining effective teachers may be an effective strategy for rapid and sustained turnaround.

In a school-level analysis in California, Dee (2012) found using a fuzzy regression discontinuity design that SIG reforms were associated with increases in school performance. In San Francisco Unified School District, a study using a difference-in-differences design found significant positive effects of SIG on student achievement, declines in unexcused absences, increased retention of high value-added teachers, and an increase in teacher-reported professional reports (Sun, Penner, & Loeb, 2017). A comparative interrupted time series analysis of a school turnaround effort in Los Angeles Unified School District (LAUSD) found that students enrolled in schools undergoing turnaround made significant gains in ELA in each of the first two years of the reform (Strunk, Marsh, Hashim, Bush-Mecenas, & Weinstein, 2016). Using an RD design, an evaluation of the Ohio SIG program found positive effects on reading and math achievement in turnaround schools (Carlson & Lavertu, 2018).

These recent successes highlight the potential for school improvement strategies to be effective when they address the root causes of failure rather than its symptoms. For example, in both Massachusetts and Tennessee, turnaround efforts focused on recruiting and retaining effective teachers (Henry et al., 2020; Papay, 2015; Zimmer et al., 2017). A successful turnaround model in Los Angeles Unified School District (LAUSD) focused on supporting school staff in implementing comprehensive, coherent reform plans that were targeted to address community-specific challenges, while success in subsequent cohorts served under the model appeared to be undermined in part by high teacher turnover (Strunk et al., 2016).

School performance and decline

Few studies have examined the role of specific dimensions of schooling on school decline. A sparse literature has examined the phenomenon of school decline, providing a guide to understanding the trends that lead to decline and the mechanisms that may contribute to a drop into low performance. However, this literature base is limited to an early qualitative study of school improvement and decline and a small subset of more recent studies that focus mostly on the process of test score decline and a few demographic variables.

In this section, I begin by describing existing studies and literature on school decline, and then conclude with a broader literature on school performance to illustrate how other school-level factors may contribute to a drop into low performance. Understanding school decline is a relevant precursor to understanding low performance because decline precedes failure (Murphy & Meyers, 2007). The symptoms of decline are therefore potential early-warning signals that a school is heading toward failure (Meyers & Murphy, 2007; Murphy & Meyers, 2007). Prior to establishing a set of factors that may contribute to decline, it may be important to first characterize the process of decline. Little research has examined this process. However, in a pair of related papers, Hochbein (2011, 2012b) provided a framework for examining school decline. In particular, he considered decline in two ways. These relational definitions of decline correspond to how a school is judged to be in decline, where individual decline is a within-school drop in performance and relational decline is a drop in performance relative to other schools in the state (Hochbein, 2011). Federal policy calls for states to classify schools using the latter approach, with low performance relative to other schools in the state. In a supplemental study more formally testing the presence of absolute, relational, and “crossing-the-line” decline—in which a school drops from above average to below average in a five-year period—in Virginia elementary schools, Hochbein (2012a) found that individual, relational, and cross-the-line decline occurred continuously over time with intermittent, one-year improvements during the downturn.

An early study of school improvement and decline used qualitative methods to examine the role of teacher and principal perceptions in school decline and improvement in eight Michigan elementary schools in 1974 through 1976 (Brookover & Lezotte, 1979). The authors found that staff in declining schools reported low levels of student ability, reported lower expectations for students to graduate high school or attend college, and were less likely to take responsibility for student success. Reflecting on these findings, Duke and Hochbein (2008) suggested that school decline research ought to systematically examine the conditions that provoke school decline. They posited that external factors may include in-migration of at-risk students, change in resources, and new mandates, while internal factors may include principal turnover, teacher turnover, and new school-level policies. A subsequent study by the authors examined only the influence of student demographic changes on decline, focusing on economically disadvantaged rates, students with disabilities, and limited English proficient students. This study, which examined student composition changes in elementary schools, found that schools with increasing rates of economically disadvantaged students experienced decreased proficiency rates in fifth-grade reading (Hochbein & Duke, 2011). The authors did not examine the influence of other factors they raised in their earlier study or school climate measures such as staff expectations and responsibility for student success.

Taken together, these studies motivate the need to better understand the mechanisms of school decline—in particular among the lowest performing schools—while also demonstrating the dearth of literature that has systematically assessed the role of school-level factors in school decline. Drawing largely from descriptive research on low-performing schools, Meyers and Murphy (2007) argued that successful turnaround requires a thorough understanding of the unique set of internal and external conditions likely to contribute to each school's low performance. A broader literature on school performance points toward the set of three dimensions of school performance described above. While this literature does not necessarily speak to school decline and failure, it informs a set of potential mechanisms of school decline.

Students. Existing research has found that changes in school demographics were associated with school decline (Hochbein & Duke, 2011). While a limitation of this research is its sole focus on school demographics as a predictor of decline, the findings underscore the importance of student-level challenges to a school's ability to effectively educate students. More generally, low-performing schools serve disproportionately high numbers of minority students and students in poverty, and students that are less prepared as they begin schooling (Meyers & Murphy, 2007; Mintrop & MacLellan, 2002). A wealth of evidence demonstrates that external conditions such as student demographics and socioeconomics influence internal conditions such as teacher quality and mobility. For example, the teacher labor market for low-performing schools is different from the market for schools with fewer challenges, with high quality teachers sorting into more affluent schools with lower student minority populations (see, e.g., Jackson, 2009). However, survey research has found that teachers value salaries and working conditions such as administrative support, class size, and facilities more than particular student characteristics (Horng, 2009; Johnson, Kraft, & Papay, 2012; Viano, Pham, Henry, Kho, & Zimmer, 2020)—suggesting a viable path forward for high poverty and minority schools trying to recruit and retain high quality teachers.

Teachers and leaders. Low-performing schools tend to rely more on teachers with lower skills and less experience than higher performing schools (Lankford, Loeb, & Wyckoff, 2002; Meyers & Murphy, 2007). While a large literature demonstrates that teachers are critical to student success (e.g., Chetty, Friedman, & Rockoff, 2013; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004), lower performing as well as higher poverty and minority schools have less effective teachers (e.g., Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2013). Low-performing schools tend to rely more on novice teachers, teachers with emergency credentials, and long-term substitutes than do more higher performing schools, and these teachers tend to be less effective and turn over at higher rates than their peers (Meyers & Murphy, 2007; Rivkin et al., 2005; Rockoff, 2004). Out-of-field teaching, which is also more prevalent in high-poverty and low-performing schools that struggle to fill teaching vacancies, is associated with lower effectiveness and higher turnover as well (Almy & Theokas, 2010; Ingersoll, 1996, 2001). In addition to experiencing high turnover among inexperienced and less effective teachers, low-performing schools also see higher rates of turnover among more effective teachers, who tend to transfer to other schools or leave the system altogether (Boyd, Lankford, Loeb, & Wyckoff, 2005).

Low-performing schools often struggle with ineffective leadership and high turnover among principals and assistant principals (Meyers & Murphy, 2007). While principals have a less direct effect than teachers on student outcomes, they influence student learning through indirect paths such as teacher retention, instructional leadership, and organizational structures (Grissom, 2011; Hallinger & Heck, 1996). Transformational leadership combined with shared instructional leadership—or distributed leadership—may be a critical ingredient in successful school reform (Marks & Printy, 2003).

School climate and conditions. Finally, existing research has documented a climate and conditions gap between low- and high-performing schools (Brown, Anfara, Jr., & Roney, 2004; Huang, Waxman, & Wang, 1995). In particular, a qualitative study of middle schools in the Philadelphia area found that teachers in low-performing middle schools reported lower expectations for students, a less orderly learning environment, and less collaboration and collegiality among colleagues than teachers in high-performing schools (Brown et al., 2004). A survey of teachers in urban elementary schools in a single district also found that teachers in low-performing schools reported more lower expectations, weaker relationships among staff, and more disciplinary problems than their peers in higher performing schools (Huang et al., 1995).

These disparities are germane because school climate and conditions play vital roles in school improvement and implementation of reforms (Bulach & Malone, 1994; Gregory, Henry, & Schoeny, 2007). Safety, relationships, teaching and learning, institutional environments, and the school improvement process can shape school climate to maximize student learning in a school improvement context (Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). Organizational climate—measured through survey items related to teacher commitment to the school and to professional development, teacher perceptions of job pressure, the extent to which teachers are involved in school decision making and encouraged to innovate, staff autonomy, and teacher cohesion—may contribute to a school's capacity to implement and sustain reforms (Dellar, 1998). More broadly, increased school safety and academic expectations are associated with increased gains in student achievement and decreased teacher turnover (Kraft, Marinell, & Shen-Wei Yee, 2016). Reduction in suspensions is associated with increased student achievement and decreased student absenteeism (Arcia, 2006; Hinze-Pifer & Sartain, 2018; Perry & Morris, 2014). Higher quality teacher collaboration in a school is associated with student achievement gains, and teachers improve at greater rates when they work in more collaborative school environments (Ronfeldt et al., 2015). Teachers also improve their effectiveness at faster rates in more supportive professional environments (Kraft & Papay, 2014). While correlates of achievement gains and successful turnaround may not represent the direct inverse of the causes of low performance, they may provide a proxy for understanding the mechanisms that contribute to school performance more generally. If transforming climate is a necessary step in turnaround, it is possible that schools with weaker climate are more prone to failure.

Theoretical framework

The theoretical framework for this study draws an analogy from early warning systems of high school dropout, which uses longitudinal data to flag students at risk of dropping out based on data that school systems already collect. If schools, like students, exhibit signals that they are at risk of failure, a framework that accounts for potential signals of decline would provide insight into the correlates of low performance and how those correlates contribute to a school's drop into low performance. The goal of the risk index follows the high school dropout early warning literature, which has iterated toward a feasible, replicable, and scalable approach to identifying students at risk of dropping out in order to target intervention strategies toward students who need them (Balfanz, Herzog, & Mac Iver, 2007; Cairns, Cairns, & Neckerman, 1989). Constructing a measure that represents a school's risk of becoming low performing will allow policymakers to identify at-risk schools and intervene in order to prevent low performance before it occurs. Additionally, identifying the particular dimensions that place a school at risk will provide policymakers and school support specialists with a springboard from which to launch interventions and supports designed to mitigate factors that may lead to low performance.

Beginning with the three dimensions of school performance and decline described above—students, teachers and leaders, and school climate and conditions—I characterize risk as a multidimensional construct that draws from existing literature on school decline and a broader literature around school performance. In particular, the conditions associated with low performance are related to students, teachers and leaders, or school climate and conditions. The student dimension includes variables related to enrollment, in-migration, grade retention, and over-age for grade. The teachers and leaders dimension includes variables related to student-teacher ratio, experience levels, educator mobility, experience, certification, and educator pay supplements (district-level supplements provided on top of salary schedule pay). The school climate and conditions dimension draws from variables related to how a school responds to school conditions, including disciplinary practices, spending, facilities, community support, teacher leadership, school leadership, instructional practices, and relational trust.

Variables included in these dimensions of school performance may signal school decline in two ways. First, the level of the prior year measure may augur an imminent drop into low performance—for example, a school experiencing a single year of very high teacher mobility may experience a large performance decline the following year. Second, the trend over multiple years may indicate an impending change in performance that would follow the risk variable's trend—for example, a steady decline in enrollment over multiple years may precede a performance decline. For this reason, I hypothesize that both single-year levels and three-year trends may contribute to a school's risk of low performance.

Methods

Data and sample

This study draws from 10 years of North Carolina administrative data from 2008-09 through 2017-18 and the North Carolina Teacher Working Conditions (TWC) survey that was administered biennially in even years during the same time period. The administrative database contains longitudinal data on all public school students, classes, and personnel in the state. I use school-level measures where available—such as value added and proficiency rates—and collapse student-, classroom-, and teacher-level measures to the school-by-year level. The most recent TWC response rate was 91 percent, with 98 percent of schools achieving response rates of at least 40 percent. Response rates in the four prior administrations ranged from 86 to 89 percent. The TWC includes questions on time use, facilities and resources, community support and involvement, student conduct, teacher leadership, school leadership, professional development, and instructional practices and support. A reliability and validity study found that the survey items load onto these eight factors, with Cronbach's alphas ranging from .86 to .96, providing evidence that the TWC is a valid and reliable instrument for measuring teacher working conditions (New Teacher Center, 2014). Additionally, existing research has found that two of these dimensions—leadership and distributed leadership—are significant predictors of teacher retention (Schweig, 2014), and that a higher overall working conditions composite is associated with higher student achievement even after controlling for other school-level variables that may affect school outcomes. I restrict the sample to the 1,578 regular public schools that were open for all 10 years of observation, have sufficient available data to construct a performance index, and have TWC response rates above the 40 percent threshold in each of the five years of survey administration.

Measures

School performance. Drawing from the state administrative data on school proficiency rates, growth, and other school-level measures, I construct a composite measure of school performance that aligns with ESSA requirements for meaningful differentiation of schools. State approaches to measuring school performance vary within ESSA requirements for meaningful differentiation of schools. Because this analysis strives to develop a risk index that can be replicated in states across the country, the measure of school performance is intended to reflect the modal state's approach to school differentiation rather than rely on North Carolina's definition of low performance. In particular, the measure includes proficiency, value-added as measured by North Carolina's value-added system (EVAAS), four-year cohort graduation rate for high schools, English language learner proficiency,[16] and chronic absenteeism. **Error! Reference source not found.** shows how I weight each indicator in ddition to a set of alternative weights I will use to examine the sensitivity of the weighting decisions.

Table 2-1. Components and weights of ESSA-compliant school performance measure

	High schools			Elementary and middle schools		
	Main	Alt 1 (higher proficiency)	Alt 2 (higher growth)	Main	Alt 1 (higher proficiency)	Alt 2 (higher growth)
Proficiency rate	0.3	0.45	0.15	0.35	0.6	0.15
EVAAS	0.15	0	0.3	0.4	0.15	0.6
Graduation rate	0.2	0.2	0.2			
EL proficiency [1]	0.1	0.1	0.1	0.1	0.1	0.1
Chronic absenteeism	0.25	0.25	0.25	0.15	0.15	0.15

[1] I do not have access to data from the English language proficiency exam taken by English learners. As a proxy, I use the percent of students classified as ELs who scored at proficient or above on the EOG reading (grades 3-8) or English 2 EOC (grade 10). As an alternative or robustness check, I can also construct a measure of the percentage of ELs in year *t* who were no longer classified as EL in year *t+1*.

[16] Because I do not have access to the EL proficiency exam scores over time, I use as a proxy for this variable the percentage of students identified as English learners who score proficient or above on the EOG or EOC reading or English 2 exam, respectively.

Correlates of low performance. Variables I test as correlates of low performance are provided in Table 2-2. While I have placed each of these variables into one of the three school risk dimensions, some could arguably fall into more than one dimension. I ultimately attempt to assign variables to dimensions based on an exploratory factor analysis prior to running a confirmatory model. Indicators hypothesized as part of the student dimension include variables related to student preparedness and success, in-migration, and enrollment changes over time. I exclude student absenteeism-related variables because chronic absenteeism is part of the measure of low performance. Indicators in the teachers and leaders dimension deal with mobility, preparation, experience, and certification. Indicators in the school climate and conditions dimension include variables related to student behavior and discipline, per pupil expenditures, and teacher perceptions of climate as measured by the Teacher Working Conditions survey. Across each of these variables, I include the one-year level and the three-year linear trend of the school because the level and trend may independently contribute to a school's risk of low performance.

Table 2-2. Variables Tested As Predictors of Low Performance

Variable	Source[1]	Description and definition
Student preparedness		
Student enrollment	*a*	Average daily membership. A drop in enrollment may signal a decline in school performance.
Minority percent	*a*	Percent of total students in the school who are racial or ethnic minorities. High rates of minority students or an increase in minority percent may signal a decline in school performance.
Black percent	*a*	Percent of total students in the school who are black. High rates of black students or an increase in black percent may signal a decline in school performance.
Hispanic percent	*a*	Percent of total students in the school who are Hispanic. High rates of Hispanic students or an increase in

Variable	Source[1]	Description and definition
		Hispanic percent may signal a decline in school performance.
Nonstructural in-migration	*a*	Percent of enrolled students who transferred into the school outside of the feeder pattern, defined as transferring while in a grade level that is less than or equal to the highest grade level offered in the sender school. High or increasing rates of nonstructural in-migration may signal a decline in school performance.
Within-year in-migration	*a*	Percent of enrolled students who transferred into the school during the school year. High or increasing rates of nonstructural in-migration may signal a decline in school performance.
Low performing in-migration	*a*	Percent of total transfers in (structural and nonstructural) who scored below proficient on any exam in their prior school year. High or increasing rates of nonstructural in-migration may signal a decline in school performance.
Retained in-migration	*a*	Percent of total transfers in (structural and nonstructural) who were retained in the prior year's grade in the year of transfer. For between-year transfers, this reflects the observed school's decision to retain the student. For within-year transfers, it reflects the prior school's decision to retain the student. In both cases, high or increasing rates of students transferring in who are repeating a grade reflect lower levels of student preparedness, and may signal a decline in school performance.
Economically disadvantaged in-migration	*a*	Percent of total transfers in (structural and nonstructural) who are classified as economically disadvantaged. High or increasing rates of economically disadvantaged students may signal a decline in school performance.
Disabled in-migration	*a*	Percent of total transfers in (structural and nonstructural) with a disability at their prior school. High or increasing rates of disabled students transferring in may signal a decline in school performance.
Grade retention percent	*a*	Percent of enrolled students who were retained in the same grade from the previous year. High or increasing rates of grade retention may signal a decline in school performance, though it may also reflect higher expectations of the school and therefore signal higher school performance.
Over-age percent	*a*	Percent of enrolled students who are over-age for their grade level. High or increasing rates of students who are over-age for their grade level may signal a decline in school performance.
Teachers and leaders		

Variable	Source[1]	Description and definition
Student-teacher ratio	*a*	Student-teacher ratio as defined by administrative data. A high or increasing student-teacher ratio may signal a decline in school performance.
Novice teacher percent	*a*	Percent of teachers with 3 or fewer years of teaching experience. High or increasing rates of novice teachers may signal a decline in school performance.
First-year teacher percent	*a*	Percent of teachers with less than 1 year of teaching experience. High or increasing rates of first-year teachers may signal a decline in school performance.
Alternative entry percent	*a*	Percent of teachers certified through alternative entry other than TFA. Teachers certified through alternative entry pathways other than TFA are more likely to turn over than other teachers. High or increasing rates of alternative entry teachers may signal a decline in school performance.
TFA percent	*a*	Percent of teachers certified through TFA. High or increasing reliance on TFA may signal a decline in school performance, although if TFA teachers are replacing long-term substitutes or teachers certified through other alternative pathways they may signal higher school performance.
Mean district teacher pay supplement	*a*	Mean teacher supplement for district. Reverse code. Low or decreasing teacher pay may signal a decline in school performance.
Mean district principal pay supplement	*a*	Mean principal supplement for district. Reverse code. Low or decreasing principal pay may signal a decline in school performance.
Teacher turnover	*a*	Percent of teachers who left the school during or at the end of the school year. High or increasing rates of teacher turnover may signal a decline in school performance unless the turnover involves strategically replacing the lowest performing teachers with the higher performing ones.
Experienced teacher turnover	*a*	Percent of teachers with more than 3 years of experience who turned over. High or increasing rates of experienced teacher turnover may signal a decline in school performance unless the turnover involves strategically replacing the lowest performing teachers with the higher performing ones.
Median teaching experience	*a*	Median years of experience of all teachers. Reverse code. Low or decreasing teacher experience may signal a decline in school performance.
Principal turnovers in last 3 years	*a*	Number of principal turnovers in year t, t-1, and t-2. More principal churn may signal a decline in school performance.

Variable	Source[1]	Description and definition
Principal experience	*a*	Years of principal experience, weighted in cases of within-year turnover or multiple principals by months in principalship. Reverse code. Low or decreasing levels of principal experience may signal a decline in school performance.
Change in principal experience from prior year	*a*	Years of principal experience for principal in year *t-1* minus years of principal experience for principal in year *t*. Low or decreasing levels of principal experience may signal a decline in school performance.
School climate and conditions		
Short-term suspension rate	*a*	Short-term suspension rate as defined by state (number of suspensions <10 days per 100 students). High or increasing use of suspensions may signal a decline in school performance.
Violent acts rate	*a*	Violent acts per 1,000 students. High or increasing rates of violent acts may signal a decline in school performance.
Per pupil expenditures	*a*	Overall per pupil expenditures. Reverse code. Low or decreasing per pupil expenditures may signal a decline in school performance.
Teacher time use	*t*	Time for planning and collaboration. Reverse code. Low or decreasing planning and collaboration time may signal a decline in school performance.
Facilities and resources factor	*t*	Resource availability. Reverse code. Low or decreasing availability and quality of facilities and resources may signal a decline in school performance.
Community support and involvement factor	*t*	Community and parent involvement. Reverse code. Low or decreasing levels of community and parent involvement may signal a decline in school performance.
Student conduct	*t*	Management of student conduct and school safety. Reverse code. Low or decreasing teacher perceptions of student conduct may signal a decline in school performance.
Teacher leadership	*t*	Teacher involvement in school practices and distributed leadership. Reverse code. Low or decreasing levels of teacher leadership may signal a decline in school performance.
School leadership	*t*	School leadership's ability to create supportive school environment. Reverse code. Low or decreasing levels of teacher perceptions of school leadership may signal a decline in school performance.
Instructional practices and support	*t*	Data and support available to teachers. Reverse code. Low or decreasing teacher perceptions of instructional practices may signal a decline in school performance.

Variable	Source[1]	Description and definition
Relational trust	*t*	Teacher perceptions of trust and support among teachers, leaders, and students. Reverse code. Low or declining relational trust may signal a decline in school performance.

[1] *a*: administrative data; *t*: Teacher Working Conditions
Variables denoted as "reverse code" are reverse-coded only in factor analysis.

Finally, I use the school U.S. Census urbanicity designation to examine heterogeneity of fit by school context. The four designations are urban, suburban, town, and rural.

Empirical strategy

In this section, I describe the empirical strategies for each of four phases of the analysis. First, I describe the descriptive analysis of school performance. Then, I describe my approach for an exploratory analysis of the predictors of low performance. The third phase draws from these predictors in an attempt to develop a measurement model representing a multidimensional risk construct. Finally, the fourth phase involves validation.

Descriptive analysis of school performance. To conduct the initial descriptive analysis of school performance trajectories, I examine the number of years schools spend in the low-performing category under each of the three measures as well as the distribution of the year-to-year changes under each of the three measures. I then graphically depict school paths into and out low performance and the magnitude of year-to-year changes in performance.

Exploratory analysis. The goal of this initial analysis is to identify a set of variables for each of the three dimensions of risk that are independently and additively related to low performance as defined by the ESSA-compliant composite measure of low performance. Isolating a relatively small set of variables with strong predictive power will allow for a parsimonious set of variables that could help states to identify schools at risk of low performance, thus minimizing burden to reproduce without losing predictive power. Machine learning provides an approach to develop a sparse model with a high level of predictive capacity without overfitting. In particular, I use lasso to shrink the regression coefficients. The lasso estimator shrinks beta estimates by constraining the sum of the absolute values of the coefficients to a constant value and then truncating variables at zero (Berk, 2008; Hastie, Tibshirani, & Friedman, 2016). In other words, it uses a soft-thresholding approach to drop variables whose shrunken regression coefficients are less than zero (Hastie et al., 2016; Tibshirani, 1996). This method therefore allows me to (1) identify the strongest predictors of performance, and (2) apply a clear decision rule around which variables to carry over to the second stage and which to drop.

There are two key benefits to using statistical learning rather than least squares regression in this particular context. First, unlike least squares regression, where highly correlated variables can lead to inflated variances and Type II error, statistical learning methods are not subject to the same concerns because they iterate through model specifications and examine fit over multiple iterations rather than examining fit within a single specification. Second, least squares regression is prone to overfitting and Type I error due to idiosyncrasies of the data, leading to coefficient estimates that are sample-dependent. By contrast, statistical learning methods incorporate information about each iterative specification into the modeling process and account for that information in the modeling process. In using the lasso to shrink coefficient estimates, I minimize the probability that selected covariates have no true relationship with the outcome. While there are many available techniques for statistical learning, there is evidence that lasso makes the best use of available covariates while maximizing precision (Hastie et al., 2016).

I include both single-year level and three-year linear trend measures within each of the three dimensions of school performance to estimate models predicting performance level and performance change in the full sample of schools. I predict the continuous performance metric rather than a binary indicator for low performance because lasso with a relatively small sample has more power to predict a continuous than a binary outcome. However, I adjust the lasso tuning parameter, λ, to prioritize accurate prediction of the bottom 5 percent of schools (in the model predicting *performance level*) to ensure the selected models yield strong prediction among the lowest performing schools.

There are multiple ways to assess fit. The common approach to assessing model fit for continuous response variables is the root mean square error (RMSE). However, while the response variable is a continuous measure of school performance, I am most interested in correctly predicting the bottom 5 percent of schools in parallel with ESSA requirements for states to identify their bottom 5 percent—not necessarily the performance score of each school across the continuum of performance.[17] I therefore examine RMSE but prioritize measures of prediction accuracy intended for binary response variables. I classify the bottom 5 percent as the bottom 5 percent across all schools regardless of school level.[18] A common approach to assessing model fit for binary outcomes is accuracy rate, which is calculated as the proportion of observations predicted correctly. In the case of class imbalance, in which the response variable is disproportionately weighted toward one value, accuracy rate will overstate prediction accuracy. In this case, only 5 percent are designated as low performing. It would therefore be possible to misclassify every low-performing school and still achieve an accuracy rate of 95 percent if the model correctly classified all non-low-performing schools. Instead, I am interested in balancing the risk of false positives against false negatives. False positives are those schools that the

[17] While ESSA requires states to identify the bottom 5 percent of their Title 1 schools in particular, I use all schools in order to maximize my sample size. Additionally, although ESSA only requires states to identify Title 1, some state plans say they will identify both Title 1 and non-Title 1 schools.

[18] 40 of 50 states use this approach in their ESSA plans.

algorithm predicts to be low performing that are not low performing, while false negatives are those schools that the algorithm predicts not to be low performing that actually are low performing. I choose the criterion that maximizes accurate prediction of low-performing schools. Following Sansone (2018), who applied similar methods to high school dropout early warning systems, I use recall rate as the primary fit criterion, defined as

$$\varphi = \frac{TruePositive}{TruePositive + FalseNegative}$$

where φ represents the proportion of low-performing schools that are accurately classified.

Importantly, I cannot select the optimal penalization term on the full sample or I risk overfitting the model (Hastie et al., 2016; Tibshirani, 1996). I therefore identify the optimal λ using bootstrapping. Specifically, I iterate through 41 values of lambda ranging from 10^1 to 10^5, sample 80 percent of the sample 1,000 times for each λ, and run the lasso regression for each sample. I then calculate the appropriate measure of fit (recall rate for performance level and RMSE for performance change) from each of those samples, calculate the means and standard deviations across all 1,000 samples for each λ, and select the λ with the best fit using grid search.

Schools are identified as low performing on a relative scale and the bottom 5 percent is an arbitrary cutoff defined by federal policy—there is no substantively meaningful shift in school performance that occurs when a school drops from the 6th percentile to the 5th. I therefore examine model fit for the bottom 8 and 10 percent of schools under the optimal λ for the bottom 5 percent. Eight and 10 are also arbitrary thresholds, but by showing model fit is not sensitive to a single arbitrary threshold I examine its potential generalizability.

On their own, the models identifying predictors of low performance would be informative to policymakers seeking to intervene in schools at risk of becoming low performing or at risk of a drop in performance. I attempt to develop a measurement model using the sets of variables providing the best model fit (as defined by φ) in the exploratory analysis. A multidimensional risk measure drawing from these variables could provide additional information to policymakers identifying schools for particular types of supports, and to researchers examining heterogeneous effects of interventions. The goal of the next stage is to construct this multidimensional measure.

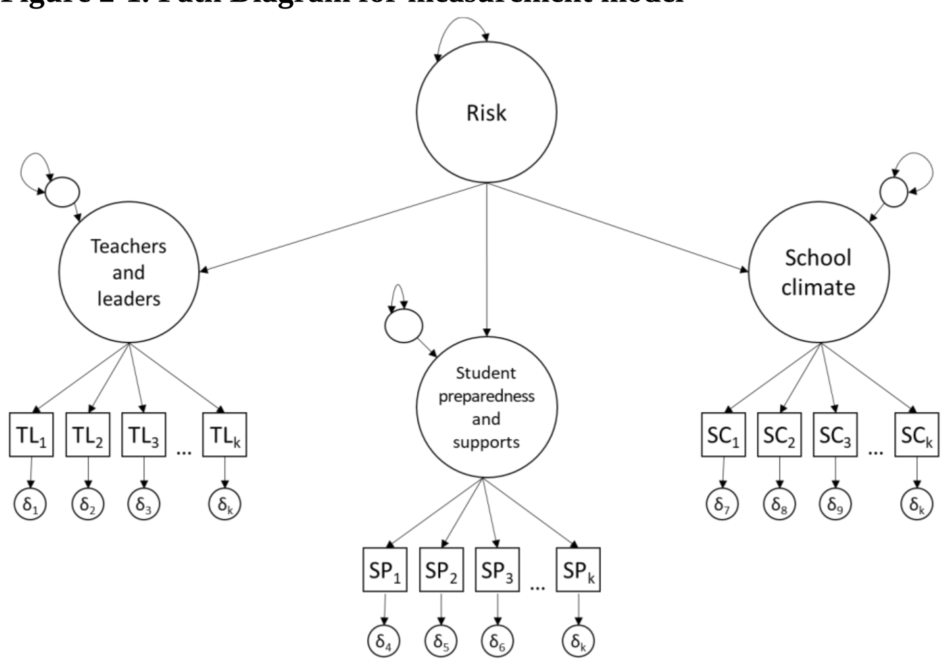
Measurement model. This stage involves developing a measurement model to construct latent variables for each of the three dimensions of risk and an overall risk factor. I begin by splitting the 1,578 schools into two random samples, stratifying on school level. The first sample, with 788 schools observed for eight years each, is intended as the calibration sample. The second sample, with 790 schools, was intended as the validation sample, though I do not undertake this phase because the calibration sample yielded poor fit. While theory guided the selection of variables to test in the first stage, no theory currently exists on the dimensions of school-level risk. A data-driven approach to identifying these dimensions would provide a framework from which future research can expand and theory can develop. I therefore run an exploratory factor analysis (EFA) with the variables identified in the previous stage using maximum likelihood estimation on the calibration sample.[19] I reverse-code variables that predicted performance index increases. Because I hypothesize that the multiple dimensions of risk are also correlated with one another, I use promax rotation criteria to allow for that correlation (Hendrickson & White, 1964)

I then attempt to fit two different structural equation models—one using the set of variables identified in the first stage and replicating the factor structure that emerged from the exploratory factor analysis, and one following the hypothesized factor structure shown in Figure

[19] I use the factor command with maximum likelihood estimation in Stata 15.

2-1. I use maximum likelihood estimation. After running the models, I examine model fit using two indices that reflect different aspects of fit, root mean square error of approximation (RMSEA, Steiger & Lind, 1980) and Tucker Lewis Index (TLI, Tucker & Lewis, 1973). TLI is an incremental fit index that compares the model of interest with a null model, while RMSEA provides a direct measure of fit rather than in relation to a null model. Both indices reward model parsimony, an important consideration in developing a reasonably straightforward measure that states can replicate. While the initial models based on variables identified in the first-stage exploratory analysis combined with theory about which of the measured variables will have correlated residuals, I make modifications as necessary to achieve convergence and improve model fit while also ensuring the factor will ultimately represent the underlying dimension of interest (Cole, Ciesla, & Steiger, 2007).

Figure 2-1. Path Diagram for measurement model



NOTE: Diagram shows general structure of Stage 2 measurement model. Manifest variables (represented by boxes) will be determined in Stage 1. Latent constructs represented by large circles. Small circles designated with deltas are residuals. Residuals will be correlated as needed.

Critical to examination of model fit is the statistical power to reject an incorrect model. An RMSEA power analysis using the approach proposed by MacCallum, Browne, & Sugawara (1996) and a null RMSEA of .05 show I have power greater than .99 to reject a close fitting model (RMSEA>.10), assuming an alpha level of .05 and a sample size of 7,101 school-by-year observations.

Prediction. Because the data do conform to a measurement model with acceptable fit, I use the manifest variables for prediction instead of the proposed latent constructs. Specifically, using the model developed in the exploratory analysis, I assess model fit overall and then by urbanicity to examine overall predictive power and heterogeneity by school context. I calculate recall and accuracy rate for the bottom 5 percent, bottom 8 percent, and bottom 10 percent of schools. I also examine RMSE and R^2 for the full model.

Additionally, to examine the decay of predictions over multiple years, I run models predicting performance in year *t+2* and *t+3*. By extending the number of years in which I predict low performance, I aim to hone in on the true predictors of low performance and maximize the lead time of the early warning index. However, as the number of years included in the low-performing outcome increases, the number of years I can use as predictors will decrease.

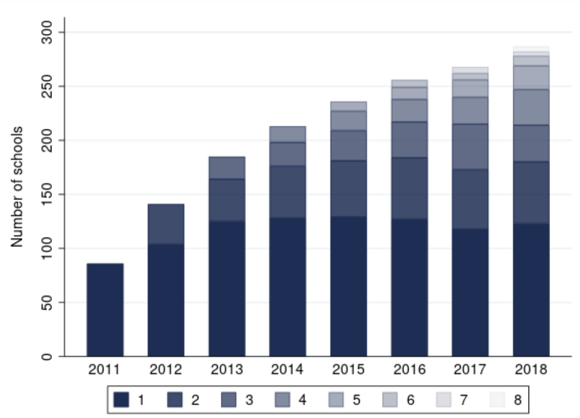
Results

School performance

Before identifying the predictors of low performance, it is helpful to understand the phenomenon of school failure. As Hochbein (2012a) found in an attempt to model school decline, there is no singular path into low performance shared across all schools. Instead, schools follow different paths into the bottom 5 percent, with some making large drops and others dipping below the arbitrary 5 percent threshold after hovering just above it in prior years. The composition of the bottom 5 percent changes from year to year as well. Few schools remain in the bottom 5 percent year after year, and many drop for just a single year. In this section, I describe the bottom 5 percent of schools in five ways. I begin by showing how often schools would be classified as low performing over an eight-year period from 2010-11 through 2017-18 under the ESSA-compliant performance index. Second, I provide context on where schools fell in the distribution of school performance in the year before becoming low performing. Third, I show the one-year performance change in schools across the performance spectrum. Fourth, I describe the year-to-year change in performance among the lowest performing schools. Finally, I provide a visual representation of the performance trajectories for a random subset of low-performing schools. Across most of these depictions of the lowest performing schools, I show descriptive statistics for the main performance index and the first two alternative indices laid out in Table 2-1.

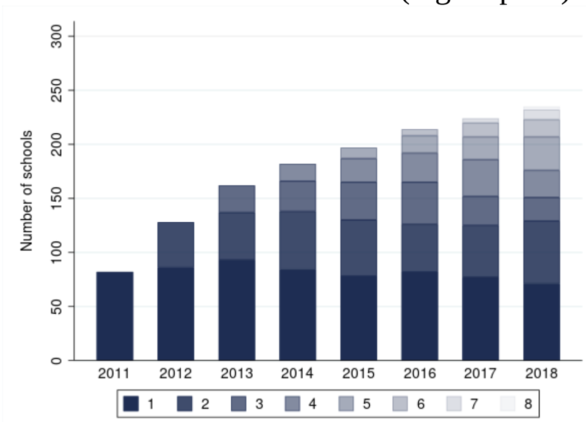
Figure 2-2. Bottom 5 percent of schools in North Carolina by year since Race to the Top

Panel A: Main index



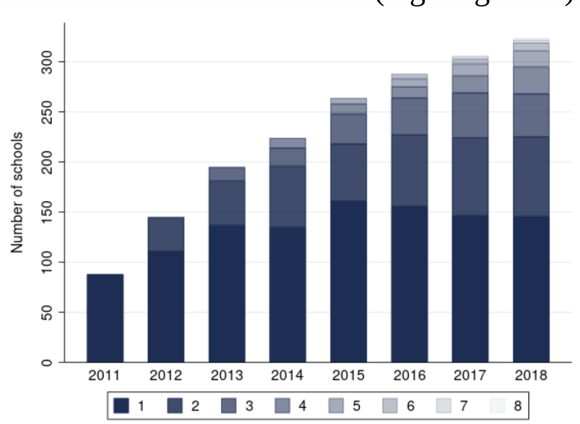
N=230 schools that were in the bottom 5 percent for at least 1 year during the eight-year span from 2010-11 through 2017-18.

Panel B: Alternative 1 index (higher prof.)



N=186 schools that were in the bottom 5 percent for at least 1 year during the eight-year span from 2010-11 through 2017-18.

Panel C: Alternative 2 index (higher growth)



N=262 schools that were in the bottom 5 percent for at least 1 year during the eight-year span from 2010-11 through 2017-18.

Colors represent number of years in bottom 5 percent, where darkest blue denotes first year in bottom 5 percent during the observed year and lightest blue represents 8th year. The measure includes proficiency, EVAAS, graduation rate for high schools, chronic absenteeism, and EL proficiency. Sample restricted to schools open all eight years of observation.

Figure 2-2 and Table 2-3 show that new schools fall into the bottom 5 percent each year, with the modal school classified as low-performing for just one year in a seven-year period. Specifically, Figure 2-2 shows the number of schools in the bottom 5 percent on the main ESSA-compliant performance measure at any point from 2011 through 2018. The bar height represents the number of schools that were in the bottom 5 percent of schools at any point between 2011 and the year shown, while the shading denotes the number of years in the bottom 5 percent during that time period. The index weighting growth more than proficiency would yield more schools spending at least one year in the bottom 5 percent, while the index weighting proficiency more would produce fewer schools identified as low performing during the eight-year period. Using the main index, nearly 60 percent of schools that dropped into the bottom 5 percent at some time from 2010-11 through 2017-18 were low performing for at least two of those years. By the final year of the eight-year period, about 14 percent of schools would have been classified as low performing at least once under the main performance index.

Table 2-3. Number of years in the bottom 5%, 2010-11 through 2016-17

Panel A: Main index

	0	1	2	3	4	5	6	7	8
2011	95.8	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012	93.2	5.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0
2013	91.0	6.1	1.9	1.0	0.0	0.0	0.0	0.0	0.0
2014	89.7	6.2	2.3	1.1	0.7	0.0	0.0	0.0	0.0
2015	88.6	6.3	2.5	1.4	0.9	0.4	0.0	0.0	0.0
2016	87.6	6.2	2.8	1.6	1.0	0.5	0.3	0.0	0.0
2017	87.0	5.7	2.7	2.0	1.2	0.8	0.3	0.3	0.0
2018	86.1	6.0	2.8	1.6	1.6	1.1	0.4	0.2	0.2
N	16504								

Panel B: Alternative 1 index (higher proficiency)

	0	1	2	3	4	5	6	7	8
2011	96.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012	93.8	4.2	2.0	0.0	0.0	0.0	0.0	0.0	0.0
2013	92.1	4.5	2.1	1.2	0.0	0.0	0.0	0.0	0.0
2014	91.2	4.1	2.6	1.4	0.8	0.0	0.0	0.0	0.0
2015	90.5	3.8	2.5	1.7	1.1	0.5	0.0	0.0	0.0
2016	89.6	4.0	2.1	1.9	1.3	0.8	0.3	0.0	0.0
2017	89.1	3.7	2.3	1.3	1.6	1.0	0.6	0.2	0.0
2018	88.6	3.4	2.8	1.1	1.2	1.5	0.8	0.4	0.1
N	16504								

Panel C: Alternative 2 index (higher growth)

	0	1	2	3	4	5	6	7	8
2011	95.7	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2012	93.0	5.4	1.6	0.0	0.0	0.0	0.0	0.0	0.0
2013	90.5	6.6	2.1	0.7	0.0	0.0	0.0	0.0	0.0
2014	89.1	6.5	3.0	0.9	0.5	0.0	0.0	0.0	0.0
2015	87.2	7.8	2.8	1.5	0.5	0.3	0.0	0.0	0.0
2016	86.0	7.6	3.4	1.8	0.5	0.4	0.2	0.0	0.0
2017	85.2	7.1	3.7	2.2	0.8	0.6	0.2	0.1	0.0
2018	84.3	7.1	3.8	2.1	1.3	0.8	0.4	0.1	0.1
N	16504								

NOTE: Sample restricted to schools open all eight years from 2010-11 through 2017-18.

While about half of schools classified as low performing in a given year were also low performing the prior year, schools dropped into the bottom 5 percent from throughout the prior year distribution of schools. The median school in the bottom 5 percent in a given year was at or near the bottom 5 percent the prior year across all measures of school performance. Still, Table 2-4 shows that low-performing schools came from across the spectrum of school performance the prior year. The bottom 5 percent of schools was more stable in indices placing more weight on proficiency and less stable in indices placing more weight on growth. Under the alternative index weighting growth more heavily, schools make larger year-to-year drops into low performance.

Table 2-4. Prior year percentile of bottom 5 percent, 2011-12 through 2017-18

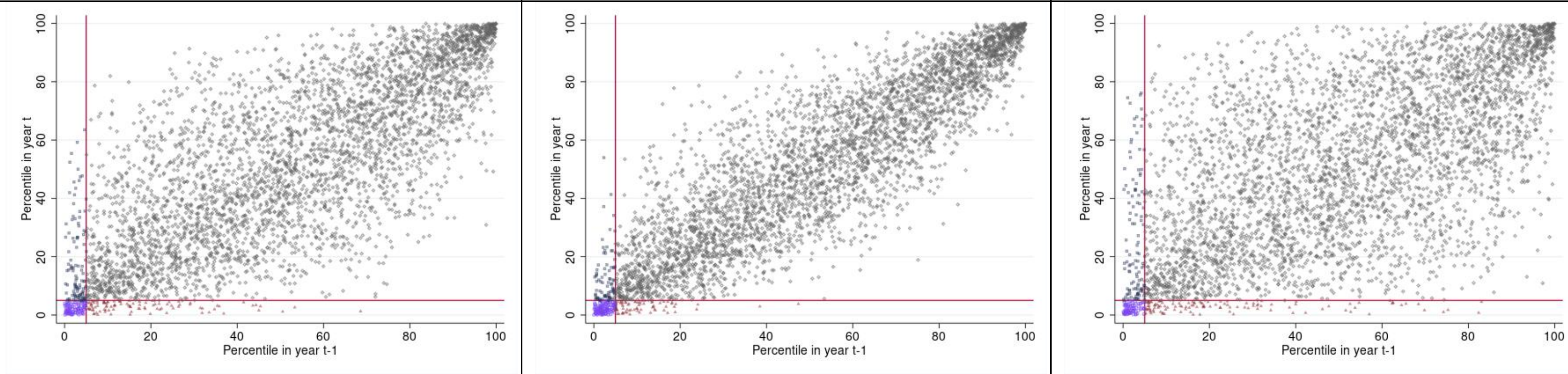
	Min	Median	Mean	Max
Main	0.05	5.34	10.26	72.3
Alt 1 (higher proficiency)	0.05	4.16	6.22	54.9
Alt 2 (higher growth)	0.05	6.93	14.38	83.9

Figure 2-3 shows that the year-to-year change in performance varies depending on choice of performance index. The alternative index that weights growth more heavily than proficiency (Panel C) yields a set of low-performing schools with more variation in prior year performance;

using this index, fewer schools experience consecutive years in the bottom 5 percent (about 40%, reflected in the lower left quadrant) and a small subset even drop from the top quartile of schools to the bottom 5 percent. The Pearson correlation between performance in year *t* and performance in *t*-1 in this index is .64, substantially lower than the .78 correlation using the main index that places less weight on growth. Meanwhile, the alternative index that weights proficiency more highly (Panel B) yields more stability in the bottom 5 percent, with about 60% of low-performing schools in a given year having been low-performing the prior year, and a year-to-year correlation of nearly .9. A state's choice of weighting is therefore likely to have implications for the stability of the CSI school list, and the likelihood of a school remaining on the list long enough to be designated as persistently low performing and targeted for more rigorous intervention.

Figure 2-3. Prior year performance of bottom 5 percent of schools

Panel A: Main index	*Panel B: Alternative 1 index (higher proficiency)*	*Panel C: Alternative 2 index (higher growth)*



r=.779	*r*=.898	*r*=.644

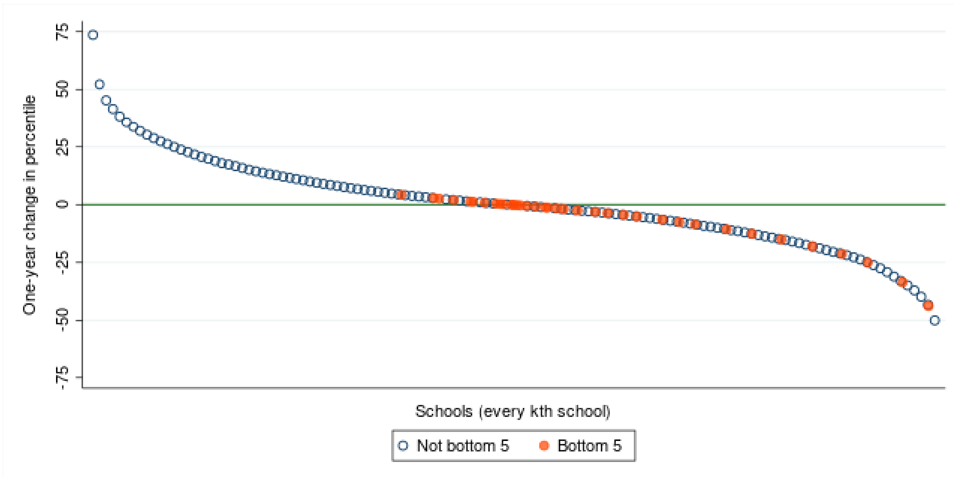
NOTE: Markers represent a random sample of 1/3 of schools per year. Red lines delineate 5th percentile. Schools in the bottom 5 percent in both years are represented as purple circles and contained in the lower left box. Schools in the bottom 5 percent in year t but not in the bottom 5 percent in t-1 are represented as red triangles to the left of the vertical red line and below the horizontal red line. Schools in the bottom 5 percent in year t-1 but not year t are represented as blue squares to the left of the vertical red line and above the horizontal red line.

While the distribution of prior performance of low-performing schools is relevant to understanding the trajectories into the bottom 5 percent, the bottom 5 percent is still an arbitrary threshold set by federal policy. To that end, the year-to-year change in school performance across the full spectrum of school performance may speak to the broader phenomenon of school improvement and decline more completely. Figure 2-4 shows the one-year change in performance across the full range of schools in the state, regardless of whether they are low performing. Specifically, it shows the size of the one-year performance change as calculated by each of the three indices. Schools overlapping the horizontal green zero line made no change from one year to the next, schools above the green line improved relative to other schools, and schools below the green line declined. If the distribution of schools did not move at all from one year to the next, the school markers would make a flat line overlapping the green line. The flatness of the distribution of these markers therefore represents the stability of the measure; again, the alternative index weighting proficiency more heavily (Panel B) is most stable, the alternative index weighting growth more heavily (Panel C) is least stable, and the main index falls between the two.

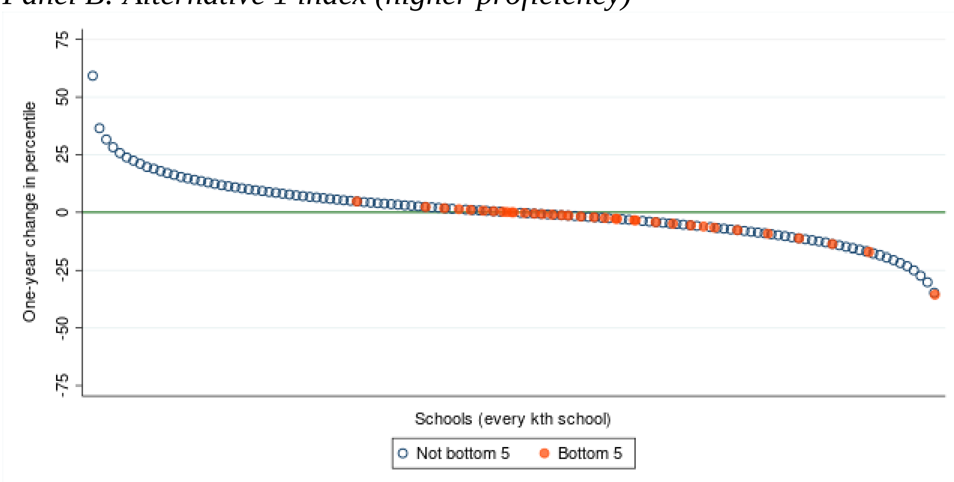
The wide range of one-year changes demonstrates that school performance using the ESSA-compliant measures can be highly variable from year to year. The clustering of many of the low-performing schools around zero suggests that absolute year-to-year changes tend to be smaller among these schools, though there are still schools that drop into the bottom 5 percent from higher performance levels. The varying steepness of the distributions by performance measure shows that the year-to-year stability of school performance is sensitive to how growth and proficiency are weighted across the spectrum of school performance.

Figure 2-4. One-year performance change of schools across the full spectrum of school performance

Panel A: Main index



Panel B: Alternative 1 index (higher proficiency)



Panel C: Alternative 2 index (higher growth)

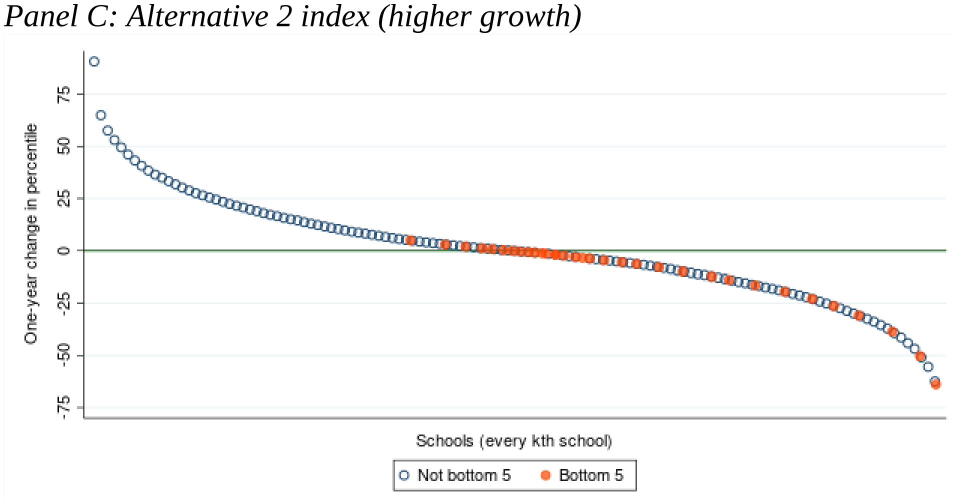
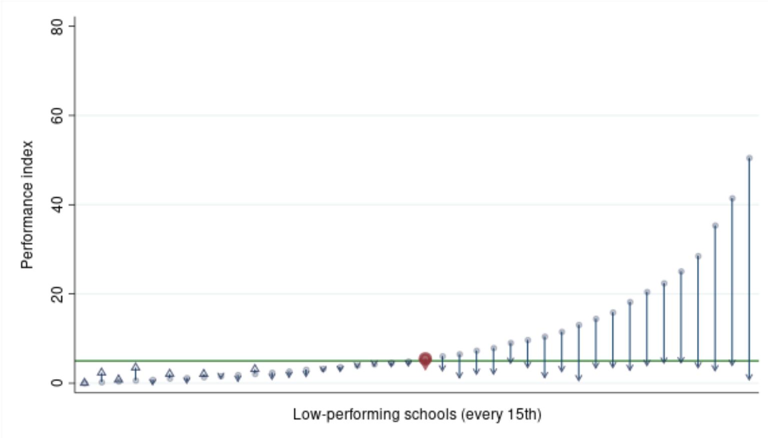


Figure 2-5 shows the range of one-year performance changes among just low-performing schools. While most low-performing schools experience a decline the year prior to the low-performing designation, some demonstrate improvement without gaining enough to move out of the bottom 5 percent. Schools make more precipitous one-year drops into the bottom 5 percent under the alternative index that weights growth more heavily (Panel C).

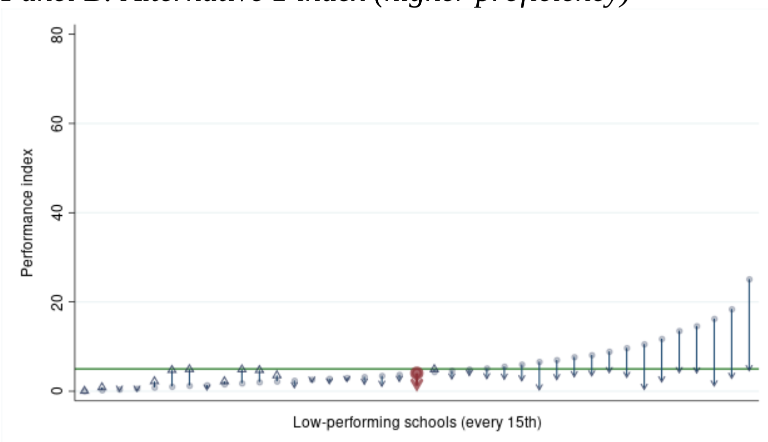
Figure 2-5. One-year performance change for bottom 5% of schools in year *t*

Panel A: Main index



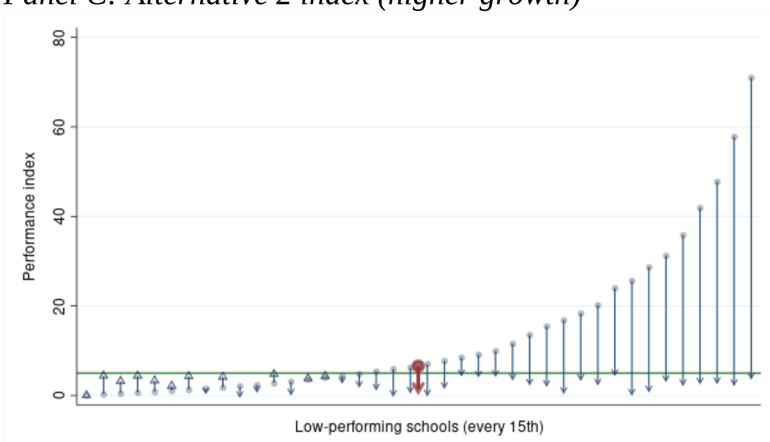
median year-to-year change: -2.91

Panel B: Alternative 1 index (higher proficiency)



median year-to-year change:-1.56

Panel C: Alternative 2 index (higher growth)



median year-to-year change: 4.52

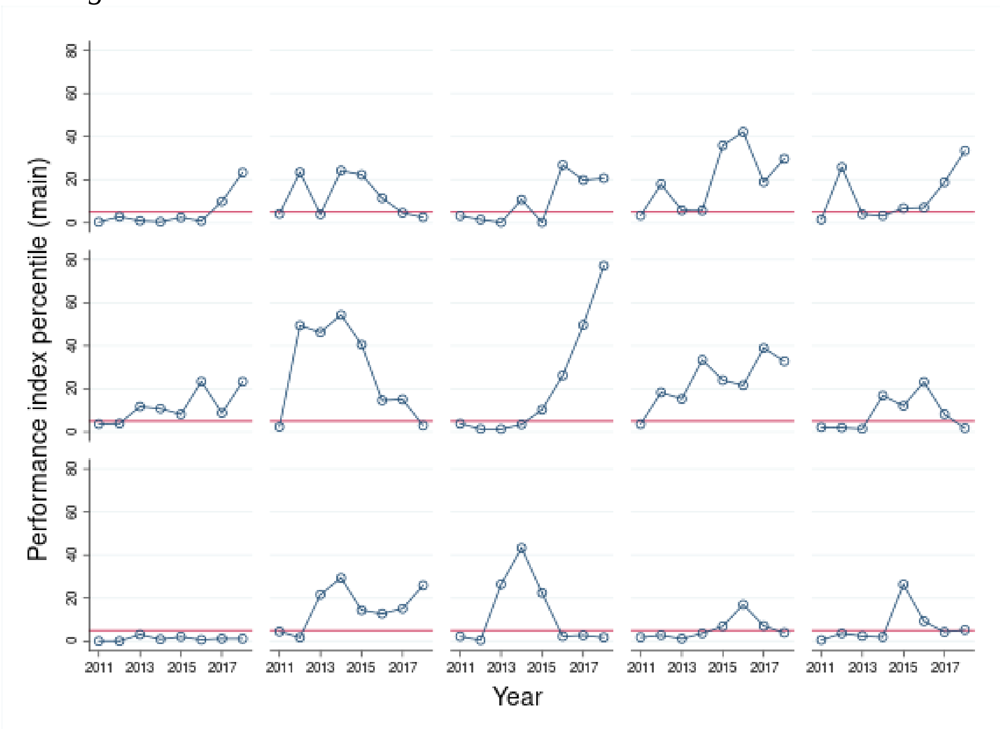
NOTE: Circle markers denote percentile in year t-1. Arrows denote percentile in year t. The vertical length of the dropline represents the size of the change from year t-1 to year t, with upward facing arrows representing schools that increased and downward facing arrows representing schools that declined. Horizontal green line demarcates the 5th percentile. Red represents median school in terms of year-to-year change.

While the above one-year performance changes provide context for understanding the extent to which schools' performance varies from year to year and how choice of weighting influences the stability of schools designated as low performing, they do not speak to the longer term trends that lead a school into or out of the bottom 5 percent. There is no single path into the bottom 5 percent. Figure 2-6 shows that while some schools remain at or near the bottom 5 percent year after year, others experience ongoing declines over time or even single-year dips from higher performance levels. Additionally, while Stuit (2010) suggested bad schools may be "immortal," these trajectories suggest there may be some schools that manage to climb from the bottom 5 percent.

Figure 2-6. Trajectories of a random sample of low-performing schools

Panel A: Main index

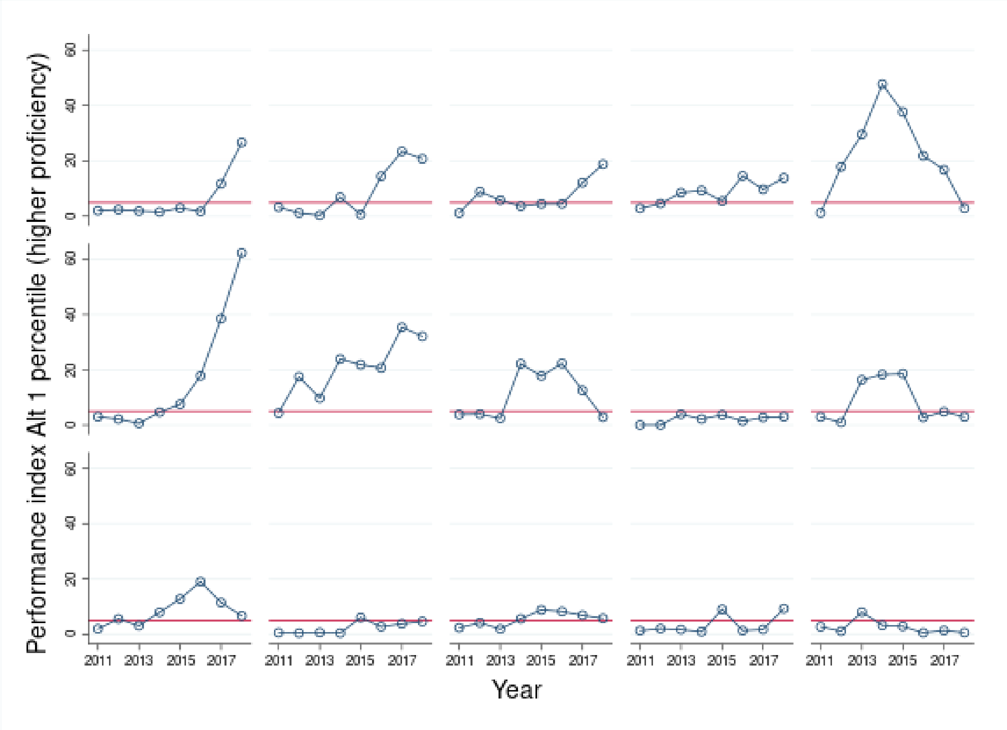
Starting in the bottom 5%



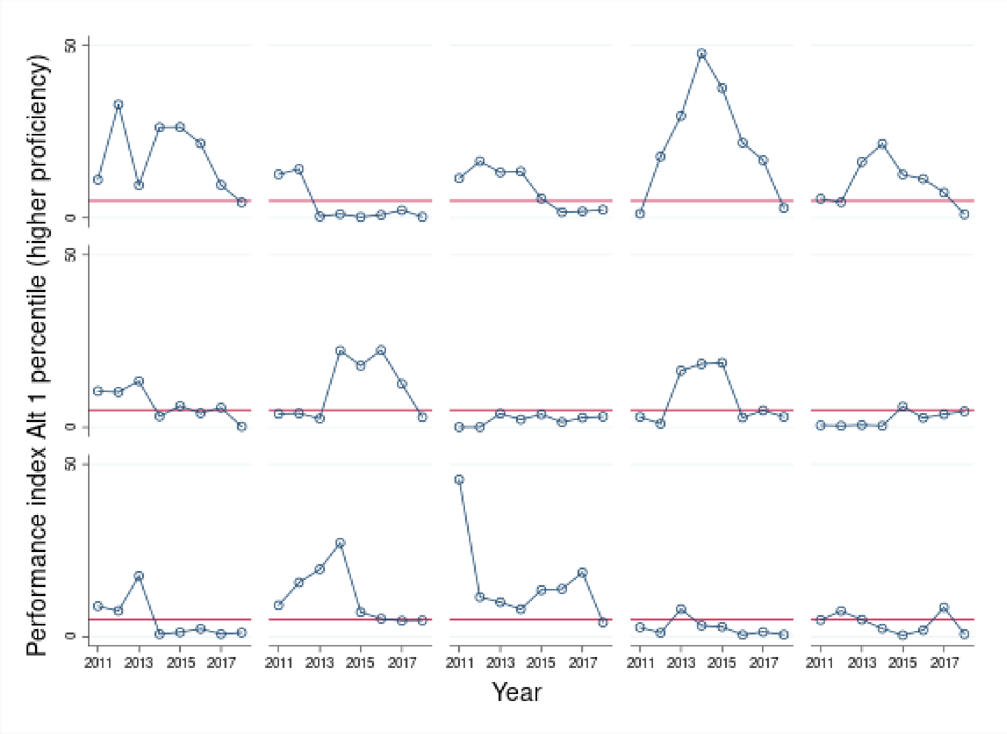
Ending in the bottom 5%

Panel B: Alternative 1 index (higher proficiency)

Starting in the bottom 5

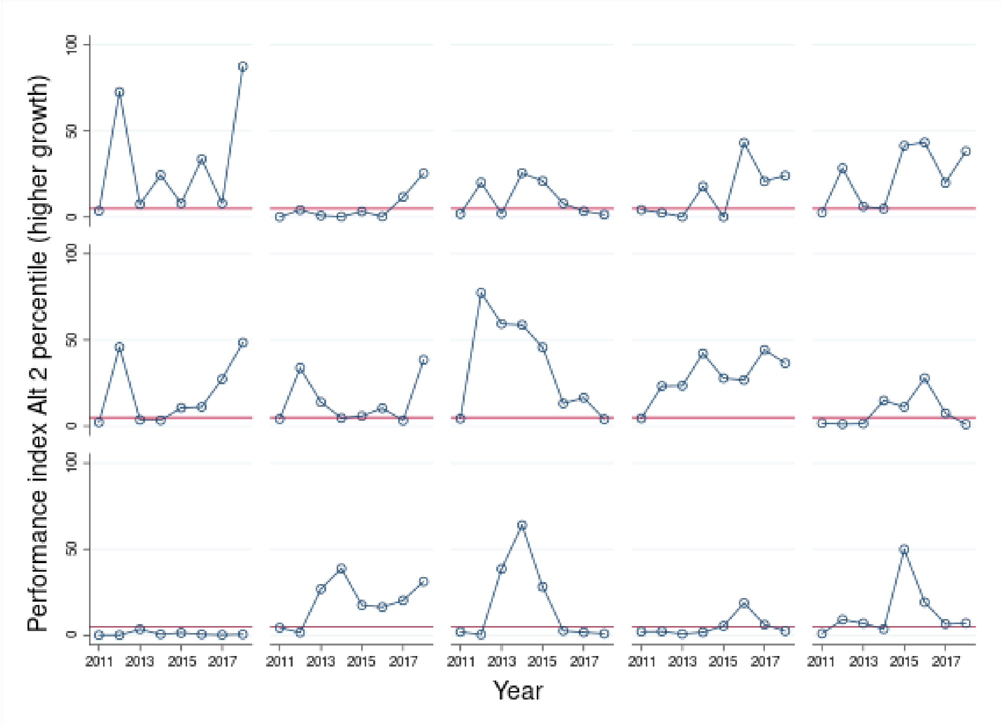


Ending in the bottom 5%

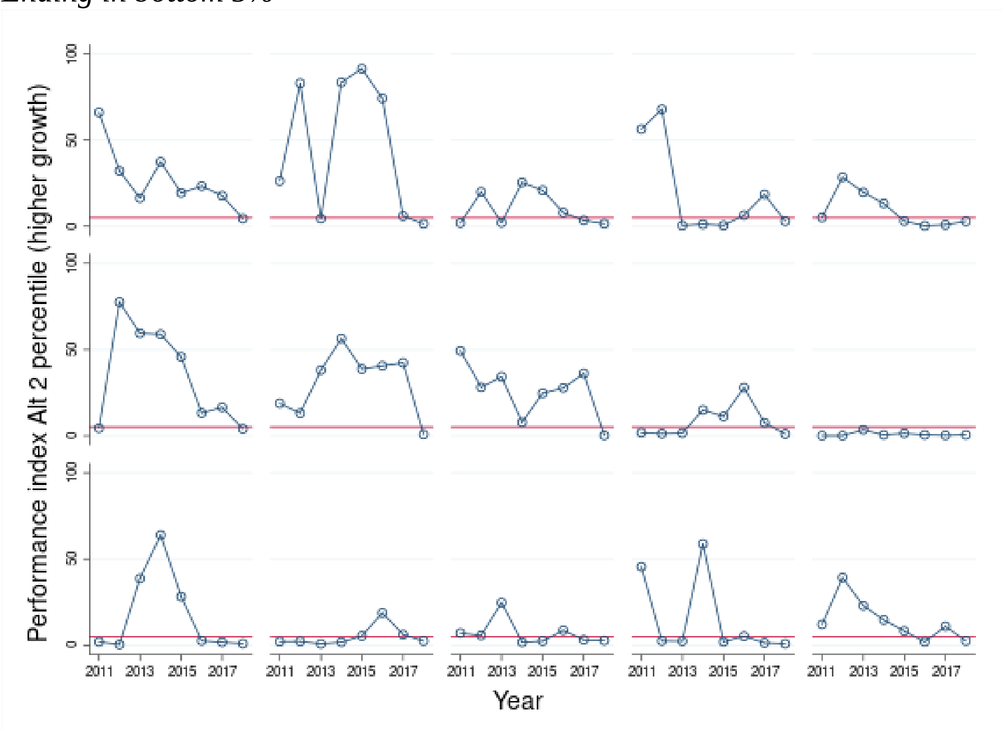


Panel C: Alternative 2 index (higher growth)

Starting in bottom 5%



Ending in bottom 5%



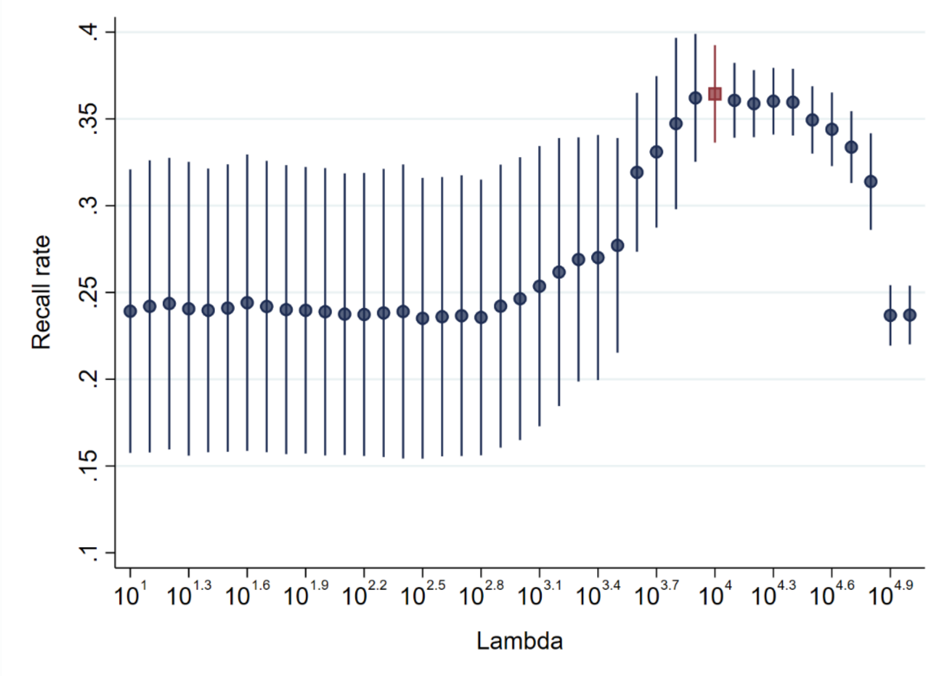
NOTE: Plots show performance trajectories 15 randomly sampled low-performing schools that were in the bottom 5% in 2011 and 15 randomly sampled low-performing schools that were in the bottom 5% in 2018. Red horizontal line demarcates the 5th percentile.

Exploratory analysis

The bootstrapping procedure to maximize predictive power identified 10^4 as the optimal value of λ for the main index performance levels outcome, yielding a mean recall rate of .364, accuracy of .940, and RMSE of 13.95. The recall rate is interpreted to mean the model accurately predicts 36.4 percent of the bottom 5 percent of schools. It is the fit statistic I want to maximize, while the accuracy rate and RMSE provide additional context about fit. The accuracy rate is interpreted to mean the model correctly predicts the status of 94 percent of all schools. The RMSE provides a measure of fit for the continuous outcome variable. As expected, RMSE is lower when using smaller values of λ because those models retain more variables that provide more predictive power overall. However, the gain in recall rate from using a larger λ outweighs the small fit improvements on the continuous measure.[20] Figure 2-7 depicts the mean recall rate over 1,000 bootstrapped samples for each value of λ from 10^1 to 10^5. The optimal λ yielded the highest mean recall rate with the smallest relative mean standard deviation over the bootstrapped samples.

[20] The full summary statistics, including RMSE, accuracy, and recall rate, for the random partition crossvalidation are available in Table A-2-1.

Figure 2-7. Mean recall rate for each λ value over 1,000 bootstrapped samples each



Markers represent mean recall rate over 1,000 bootstrapped samples for each of 41 values of λ. Spikes are 95% confidence intervals. Optimal λ in red.

A lasso regression predicting the main performance index on the full sample of schools using 10^4 as λ yielded a recall rate of .375 for the bottom 5 percent of schools, .422 for the bottom 8 percent, and .446 for the bottom 10 percent. The recall rate was highest in the model predicting the alternative index weighting proficiency more heavily and lower in the model predicting the alternative index with higher growth. Table 2-5 shows a confusion matrix, which shows the number of true positives (observed low performing and predicted low performing), false negatives (observed low performing and predicted non-low-performing), true negatives (observed non-low-performing and predicted non-low-performing) and false positives (observed non-low-performing and predicted low performing) for the model predicting the main performance index. The 189 true positives are reflected in the numerator of the recall rate. The goal of the model is to maximize this number relative to the false negatives.

Table 2-5. Confusion matrix (main index)

Actual →	Bottom 5			Bottom 8			Bottom 10		
Predicted ↓	LP	Not LP	Total	LP	Not LP	Total	LP	Not LP	Total
Low performing	189	322	511	348	731	1079	465	884	1349
Not low performing	315	10220	10535	476	12626	13102	578	12254	12832
Total	504	10542	11046	824	13357	14181	1043	13138	14181

Numbers in cells represent the number of schools falling into the true positive, true negative, false positive, and false negative categories for the bottom 5%, bottom 8%, and bottom 10% of schools.

The lasso regression for the main index retained 12 variables, shown in Table 2-6 under their hypothesized dimensions of risk. The two alternative indices retained similar variables. All retained variables represented one-year levels; no three-year trajectory variables had strong enough predictive power in the models to warrant inclusion. Of particular note is that high levels of student in-migration are not the strongest predictors of low performance; what seems to matter most is the type of in-migration—the model retained variables representing in-migration of students with disabilities and who are economically disadvantaged but did not retain overall nonstructural in-migration or even within-year in-migration. Also of note is that no direct measures of principals were retained; however, the model did retain variables on which principals have direct influence, such as school climate and community support and involvement.

Table 2-6. Predictors of low performance

	Main index	Alt 1 (higher proficiency)	Alt 2 (higher growth)
Students			
Student enrollment	X		X
Minority percent	X	X	
Black percent	X	X	X
Hispanic percent		X	
Economically disadvantaged in-migration	X	X	X
Disabled in-migration	X	X	X
Teachers			
Student-teacher ratio	X	X	X
Novice teachers	X	X	X
Alternative entry teachers	X	X	X
Mean district teacher pay supplement	X	X	
School climate and conditions			
Per pupil expenditures	X	X	X
Community support and involvement	X	X	X
Student conduct	X	X	X
Violent acts rate		X	X

Predictors identified using lasso regression with λ of 10^4.

The coefficients on the lasso regression are biased downward by design. A post-estimation OLS regression provides unshrunken coefficient estimates that provide some additional insight about the magnitude and directionality of the estimates. I present the post-lasso OLS estimates in Table 2-7 with standardized coefficients in order to compare the relative predictive power of each variable. Because the models predict the performance index, negative coefficients denote variables associated with a decrease in school performance. The variables with negative coefficients therefore represent risk factors for low performance, while the variables with positive coefficients represent mitigating factors.

Table 2-7. Post-lasso regression results predicting performance level on three performance indices (standardized coefficients)

	(1) Main index	(2) Alt 1: Higher prof	(3) Alt 2: Higher growth
Student enrollment (100s)	0.048***		0.089***
	[5.07]		[9.70]
Minority students	-0.109***	-0.037	
	[-7.12]	[-1.49]	
Black students	-0.079***	-0.144***	-0.121***
	[-5.83]	[-6.64]	[-13.47]
Hispanic students		-0.048***	
		[-3.48]	
Economically disadvantaged in-migration	-0.195***	-0.200***	-0.239***
	[-19.64]	[-20.23]	[-26.00]
Disabled in-migration	-0.048***	-0.050***	-0.043***
	[-6.44]	[-6.77]	[-5.82]
Student-teacher ratio	-0.008	-0.003	0.017
	[-0.79]	[-0.28]	[1.93]
Novice teachers	-0.076***	-0.078***	-0.083***
	[-9.68]	[-9.86]	[-10.68]
Alternative entry teachers	-0.111***	-0.101***	-0.138***
	[-12.53]	[-11.40]	[-15.73]
Teacher supplement	0.122***	0.136***	
	[13.49]	[15.40]	
Short-term suspension rate	-0.103***	-0.099***	-0.114***
	[-11.34]	[-10.31]	[-11.97]
Per pupil expenditures	-0.052***	-0.064***	
	[-5.28]	[-6.69]	
Community engagement	0.214***	0.209***	0.239***
	[17.37]	[16.75]	[19.59]
Student conduct	0.114***	0.105***	0.092***
	[10.29]	[9.59]	[8.43]

	(1) Main index	(2) Alt 1: Higher prof	(3) Alt 2: Higher growth
Violent acts rate		-0.010	-0.013
		[-1.16]	[-1.53]
Recall (B5)	0.375	0.398	0.287
Recall (B8)	0.422	0.468	0.352
Recall (B10)	0.446	0.506	0.380
Accuracy (B5)	0.942	0.945	0.933
Accuracy (B8)	0.915	0.920	0.906
Accuracy (B10)	0.897	0.906	0.887
RMSE	13.956	13.965	14.092
R^2	0.512	0.511	0.502
Obs	10526	10525	10525

Standardized coefficients. T-statistics in brackets.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The three indices yield similar results. Within the student dimension, in-migration of students with disabilities and who are economically disadvantaged were the strongest negative predictors of school performance. Within the teacher dimension, high rates of alternative entry and novice teachers were strong predictors. High levels of teacher perceptions of community engagement and student conduct were associated with increased school performance.

Measurement model

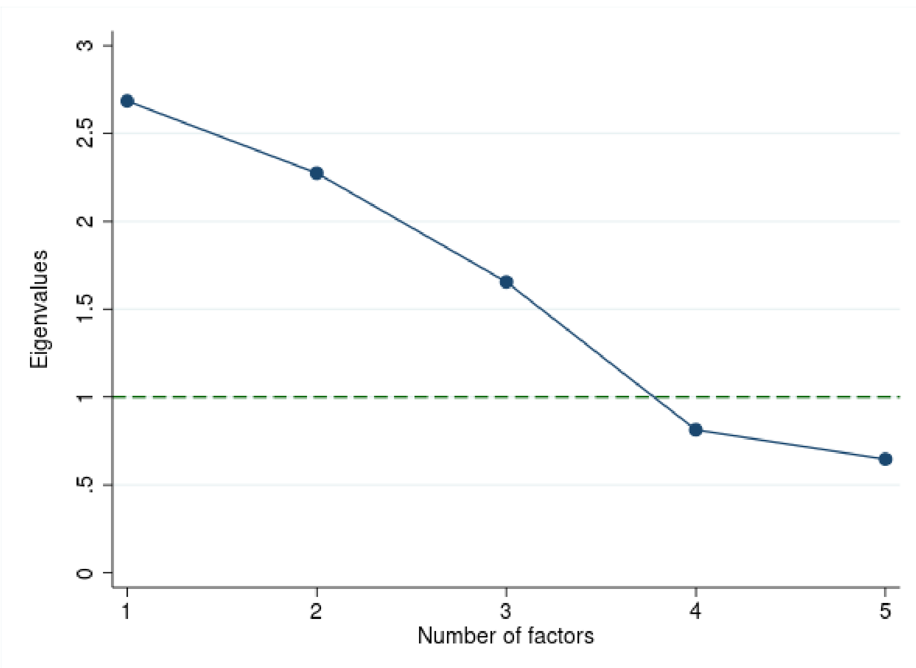
I draw from the 12 retained variables in the exploratory stage to attempt to develop a measurement model for a multidimensional measure of risk. No clear factor structure emerged using promax rotation criteria to allow for correlation between the risk dimensions. While the eigenvalues point to a three-factor solution (

Figure 2-8), model information criteria presented in Table 2-8 do not identify a clear number of factors.

Table 2-8. Information criteria by number of retained factors

	Log likelihood	df(model)	df(residual)	AIC	BIC
1	-11301.59	13	65	22629.18	22717.88
2	-6486.61	25	53	13023.22	13193.79
3	-3022.40	36	42	6116.80	6362.43
4	-1481.27	46	32	3054.54	3368.40
5	-393.59	55	23	897.17	1272.43
6	-246.90	63	15	619.80	1049.65
7	-132.25	70	8	404.49	882.10
8	-16.51	76	2	185.03	703.57

Figure 2-8. Screeplot



Screeplot presents eigenvalues for factor analysis for up to five factors.

Allowing two additional factors does not generate meaningful factors.[21] The rotated factor loadings are presented in Table 2-9. Variables for black and minority student percentage load onto one factor, while community engagement and student conduct load onto another factor—but overall, the factors explain little variation in the remaining variables.

Table 2-9. Rotated factor loadings

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Psi
Enrollment	0.026	-0.107	0.411	-0.121	0.429	0.422
Minority students	1.005	-0.013	0.010	-0.006	-0.004	0.000
Black students	0.782	-0.067	0.033	0.244	0.044	0.257
Econ. dis. in-migration	0.334	0.122	0.019	-0.118	0.699	0.332
Disabled in-migration	-0.194	-0.010	0.243	0.027	0.216	0.838
Student-teacher ratio	0.019	-0.111	-0.811	0.002	-0.069	0.281
Novice teachers	0.380	0.192	-0.063	-0.028	0.049	0.770
Alternative entry teachers	0.094	0.066	-0.042	0.631	-0.021	0.489
Teacher supp. (rev)	-0.285	0.016	-0.119	0.254	0.712	0.459
Short-term susp. rate	0.119	0.034	0.077	0.717	0.073	0.392
Per pupil expenditures	0.048	-0.017	0.895	0.084	-0.170	0.304
Comm. engagement (rev)	-0.019	1.005	0.052	-0.022	0.133	0.000
Student conduct (rev)	-0.061	0.743	-0.010	0.093	-0.157	0.382

Factor loadings from EFA with oblique varimax rotation. ψ provides the uniqueness of the variance for each variable.

A hierarchical confirmatory factor analysis as shown in Figure 2 above, which treats the three factors as a higher-order general factor of risk, did not converge. A model with three separate factors (students, teachers, and school conditions) but no overall dimension of risk did converge but revealed poor fit, as expected given the EFA results (RMSEA=.191, TLI=.510).

[21] An EFA with orthogonal rotation criteria (varimax) also did not yield meaningful factors. Rotated factor loadings using orthogonal rotation criteria are presented in Table A-2-2.

Prediction

Because the measurement model stage did not reveal a latent measure of risk, I focus prediction on the variables identified by the exploratory stage. As shown in Table 2-7 above, the recall rate for the main index is .375 for the bottom 5 percent of schools, .422 for the bottom 8 percent, and .446 for the bottom 10 percent. Table 2-10 shows recall and accuracy rates by school urbanicity. Recall rate for the bottom 5 percent is similar for urban and rural schools at about .4, though the model predicts the bottom 8 and 10 percent in urban schools more accurately than in rural schools. Very few schools with suburban and town U.S. Census designations are classified as low performing.

Table 2-10. Model fit by school urbanicity

	Urban	Suburban	Town	Rural
Recall				
bottom 5%	0.396	0.556	0.095	0.398
bottom 8%	0.517	0.381	0.154	0.414
bottom 10%	0.530	0.548	0.207	0.435
Accuracy				
bottom 5%	0.937	0.982	0.945	0.938
bottom 8%	0.903	0.962	0.888	0.917
bottom 10%	0.886	0.956	0.853	0.899
Obs				
N LPS (b5)	302	31	92	618
N	4040	1473	1370	7298

I turn next to the decay of predictions over multiple years by examining predictive accuracy for two and three years out. I find that while the model loses some predictive accuracy as it attempts to predict performance for each subsequent year out, it continues to accurately identify more than one-third of the bottom 5 percent of schools using data from two years prior and about 30 percent of the bottom 5 percent of schools using data from three years prior, as shown by the recall rates in Table 2-11. The coefficient estimates are also similar across models that predict further years out.

Table 2-11. Model predictions for year *t*+1, *t*+2, and *t*+3 (main index)

	(1) *t*+1	(2) *t*+2	(3) *t*+3
Student enrollment (100s)	0.265***	0.197***	0.127*
	(0.0523)	(0.0572)	(0.0632)
Minority students	-0.087***	-0.090***	-0.111***
	(0.0122)	(0.0133)	(0.0148)
Black students	-0.078***	-0.079***	-0.081***
	(0.0133)	(0.0146)	(0.0163)
Economically disadvantaged in-migration	-0.200***	-0.189***	-0.184***
	(0.0102)	(0.0111)	(0.0124)
Disabled in-migration	-0.170***	-0.148***	-0.153***
	(0.0264)	(0.0294)	(0.0330)
Student-teacher ratio	-0.081	0.104	0.086
	(0.1028)	(0.1216)	(0.1366)
Novice teachers	-0.139***	-0.162***	-0.185***
	(0.0144)	(0.0159)	(0.0177)
Alternative entry teachers	-0.205***	-0.226***	-0.228***
	(0.0164)	(0.0178)	(0.0195)
Teacher supplement	0.001***	0.002***	0.002***
	(0.0001)	(0.0001)	(0.0001)
Short-term suspension rate	-0.117***	-0.102***	-0.082***
	(0.0103)	(0.0113)	(0.0124)
Per pupil expenditures	-0.064***	-0.066***	-0.084***
	(0.0121)	(0.0158)	(0.0174)
Community engagement	8.353***	8.837***	8.221***
	(0.4809)	(0.5356)	(0.5909)

	(1) t+1	(2) t+2	(3) t+3
Student conduct	4.513***	2.245***	1.092*
	(0.4386)	(0.4910)	(0.5410)
Constant	80.011***	77.308***	79.717***
	(2.6000)	(3.2111)	(3.5939)
Recall (B5)	0.375	0.347	0.302
Recall (B8)	0.422	0.394	0.372
Recall (B10)	0.446	0.422	0.398
Accuracy (B5)	0.942	0.940	0.935
Accuracy (B8)	0.915	0.909	0.906
Accuracy (B10)	0.897	0.890	0.887
RMSE	13.956	14.259	14.528
R^2	0.512	0.493	0.480
Obs	10526	9022	7518

Model for year t+1 uses all outcome years from 2012 through 2018. Model for year t+2 uses 2013 on, and model for t+3 uses 2014 on. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Discussion

There are some limitations to this analysis. First, because I rely on a single state, the generalizability of these results may be limited for other states. However, North Carolina is a large, diverse state, and provides a useful context for examining low performance. Additionally, North Carolina is one of only a few states that uses value-added as a school growth measure. Future research could consider less rigorous methods for measuring growth, such as student growth percentiles. Finally, I identified low-performing schools using the school's overall rank in the distribution of schools because many states proposed that approach in their ESSA plans. However, the different formulas for meaningful differentiation of schools by school level may lead states to take a different approach and identify schools separately by level. Future research could conduct similar analyses in which the lowest performing schools are identified separately by elementary, middle, and high school rather than overall.

The leading indicators of low performance identified in this study provide context for states and districts thinking about how to identify and support low-performing schools. The strongest student-related risk factors that emerged were related to in-moving of students who may need additional resources to help them succeed. Schools with high rates of students with disabilities and economically disadvantaged students are tasked not just with providing high quality instruction, but with providing students with the supports they need to be successful in the classroom. The most detrimental teacher-related risk factors are related to teacher experience and preparation. Schools that need to rely largely on novice and alternative entry teachers because of limited local labor markets may need additional state supports to develop these teachers. Additional resources may help some schools recruit more effective teachers, and the small but significant coefficient on teacher supplement suggests that targeting funds toward higher teacher pay may help to mitigate risk factors associated with low performance. Within the school climate and conditions dimension, exclusionary discipline emerged as the primary risk factor for low performance, though this analysis cannot speak to the extent to which this risk factor captures overreliance on out-of-school suspensions versus a disorderly school environment that leads to higher use of out-of-school suspension. The models predicting the two alternative performance indices include the school violent acts rate, providing additional evidence that an unsafe school environment is a significant risk factor of school underperformance.

Two mitigating factors of school underperformance are both related to school climate and conditions. One of these factors—student conduct—adds to the evidence above. The student conduct factor, generated using the NC Teacher Working Conditions survey, asks teachers about whether students and faculty understand expectations for student conduct, whether students follow rules, whether teachers and administrators enforce rules, and whether the school environment is safe. Fostering a safe and supportive school environment with clearly established rules for student conduct may help schools to mitigate some of the risks of low performance by creating a structure that is supportive to student learning even as other risk factors may increase. Meanwhile, the community support and involvement factor, which asks teachers about parent and guardian involvement in the school, whether the school communicates with the community, and whether the parents and the broader community are supportive of teachers and the school, emerged as another mitigating factor of underperformance. This finding suggests that schools may benefit from fostering positive relationships with families and the broader community—and that states may gain some traction by targeting supports to help at-risk schools engage their communities.

The findings from this study have implications for policy around identifying and supporting low-performing schools. The current system for school turnaround, in which states identify and target the lowest 5 percent of schools in a given year, requires schools to fail before they can receive supports and resources to improve student performance. While a valid multidimensional measure of risk would have provided a means for policymakers to target supports to schools with particular challenges, the small set of predictors provides useful information about the early signals of school underperformance. The model yields a recall rate of .375 for the bottom 5 percent of schools, .422 for the bottom 8 percent, and .446 for the bottom 10 percent. While this approach does not accurately identify most low-performing schools, the approach yields a recall rate similar to the average for high school early warning indicators, where recall rates range from .044 to .969 with a mean of about .4 (Bowers, Sprott, & Taff, 2012). A recent study that applied machine learning to high school early warning systems produced recall rates of .16 to .28 using lasso methods (Sansone, 2018).

Since states expanded their longitudinal data systems under RttT, they have access to more data than ever before and the tools to store and analyze those data. In prioritizing parsimony of the risk measure, I isolate a small number of predictors that states can aim to replicate using their own data systems—either through data they already have or can begin to collect and track. This work therefore has the potential to inform state decision making around tracking, identifying, supporting, and improving low-performing schools.

Finally, by constructing ESSA-compliant index scores using different weighting schemes, I preview potential consequences of state decisions around how to construct their meaningful differentiation measure under ESSA. In particular, these decisions may have implications for the stability of the bottom 5 percent, including how many schools are identified as “persistently low performing” under ESSA.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, *25*, 95–135.

Almy, S., & Theokas, C. (2010). *Not Prepared for Class: High-Poverty Schools Continue to Have Fewer In-Field Teachers, Education Trust* (No. ED543217). Washington, D.C.: The Education Trust.

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society*, *38*, 359–369.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions. *Educational Psychologist*, *42*, 223–235.

Berk, R. A. (2008). *Statistical learning from a regression perspective*. New York: Springer.

Bonilla, S., & Dee, T. (2017). *The Effects of School Reform Under NCLB Waivers: Evidence from Focus Schools in Kentucky* (Working Paper No. 23462). National Bureau of Economic Research.

Bowers, A. J., Sprott, R., & Taff, S. A. (2012). Do We Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity, and Specificity. *The High School Journal*, *96*, 77–100.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review*, *95*, 166–171.

Brookover, W. B., & Lezotte, L. W. (1979). *Changes in School Characteristics Coincident With Changes in Student Achievement. Occasional Paper No. 17.*

Brown, K. M., Anfara Jr, V. A., & Roney, K. (2004). Student achievement in high performing, suburban middle schools and low performing, urban middle schools: Plausible explanations for the differences. *Education and Urban Society*, *36*, 428–456.

Bulach, C., & Malone, B. (1994). The Relationship of School Climate to the Implementation of School Reform. *ERS Spectrum*, *12*, 3–8.

Cairns, R. B., Cairns, B. D., & Neckerman, H. J. (1989). Early School Dropout: Configurations and Determinants. *Child Development*, *60*, 1437–1452.

Carlson, D., & Lavertu, S. (2018). School Improvement Grants in Ohio: Effects on Student Achievement and School Administration. *Educational Evaluation and Policy Analysis*, 0162373718760218.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). *Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates* (SSRN Scholarly Paper No. ID 2325796). Rochester, NY: Social Science Research Network.

Cohen, D. K., & Moffitt, S. L. (2010). *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Harvard University Press.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381–398.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement. *Education Policy Analysis Archives*, *8*, 1.

Dee, T. (2012). *School Turnarounds: Evidence from the 2009 Stimulus* (Working Paper No. 17990). National Bureau of Economic Research.

Dee, T., & Dizon-Ross, E. (2019). School Performance, Accountability, and Waiver Reforms: Evidence From Louisiana. *Educational Evaluation and Policy Analysis*, 0162373719849944.

Dellar, G. B. (1998). School Climate, School Improvement and Site-based Management. *Learning Environments Research*, *1*, 353–367.

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in Texas. *The Georgetown Public Policy Review*, 21–39.

Dougherty, S. M., & Weiner, J. M. (2017). The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance. *Educational Policy*, 0895904817719520.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., … Giffin, J. (2017). *School Improvement Grants: Implementation and Effectiveness. NCEE 2017-4013*. National Center for Education Evaluation and Regional Assistance.

Duke, D. L., & Hochbein, C. (2008). Rising to the Challenges of Studying School Decline. *Leadership and Policy in Schools*, *7*, 358–379.

Gregory, A., Henry, D. B., & Schoeny, M. E. (2007). School Climate and Implementation of a Preventive Intervention. *American Journal of Community Psychology*, *40*, 250–260.

Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, *113*, 2552–2585.

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*, 111–126.

Hallinger, P., & Heck, R. H. (1996). Reassessing the Principal's Role in School Effectiveness: A Review of Empirical Research, 1980-1995. *Educational Administration Quarterly*, *32*, 5–44.

Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2nd edition). New York, NY: Springer.

Hemelt, S. W., & Jacob, B. (2017). *Differentiated Accountability and Education Production: Evidence from NCLB Waivers* (Working Paper No. 23461). National Bureau of Economic Research.

Hemelt, S. W., & Jacob, B. A. (2018). How Does an Accountability Program that Targets Achievement Gaps Affect Student Performance? *Education Finance and Policy*, 1–68.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65–70.

Henry, G. T., & Harbatkin, E. (2019). Why Do Schools Fail? An Examination of the School-level Factors Associated with Low Performance. *Annual Conference*. Presented at the Association for Education Finance and Policy, Kansas City, MO.

Henry, G. T., Pham, L. D., Kho, A., & Zimmer, R. (2020). Peeking Into the Black Box of School Turnaround: A Formal Test of Mediators and Suppressors. *Educational Evaluation and Policy Analysis*, *42*, 232–256.

Hinze-Pifer, R., & Sartain, L. (2018). Rethinking Universal Suspension for Severe Student Behavior. *Peabody Journal of Education*, *93*, 228–243.

Hochbein, C. (2011). Overlooking the descent: Operational definition, identification, and description of school decline. *Journal of Educational Change*, *12*, 281–300.

Hochbein, C. (2012a). Downward Spirals, Boiled Frogs, and Catastrophes: Examining the Rate of School Decline. *Leadership and Policy in Schools*, *11*, 66–91.

Hochbein, C. (2012b). Relegation and Reversion: Longitudinal Analysis of School Turnaround and Decline. *Journal of Education for Students Placed at Risk (JESPAR)*, *17*, 92–107.

Hochbein, C., & Duke, D. L. (2011). Crossing the line: Examination of student demographic changes concomitant with declining academic performance in elementary schools. *School Effectiveness and School Improvement*, *22*, 87–118.

Horng, E. L. (2009). Teacher Tradeoffs: Disentangling Teachers' Preferences for Working Conditions and Student Demographics. *American Educational Research Journal*, *46*, 690–717.

Huang, S. L., Waxman, H. C., & Wang, M. C. (1995). *Comparing School-Based Environment of High- and Low-Performing Inner City Schools*. Retrieved from https://eric.ed.gov/?id=ED386503

Ingersoll, R. M. (1996). *Out-of-Field Teaching and Educational Equality. Statistical Analysis Report.* ERIC.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*, 499.

Jackson, C. K. (2009). Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation. *Journal of Labor Economics*, *27*, 213–256.

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

Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence From Panel Data. *American Educational Research Journal*, *53*, 1411–1449.

Kraft, M. A., & Papay, J. P. (2014). Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience. *Educational Evaluation and Policy Analysis*, *36*, 476–500.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, *24*, 37–62.

Leithwood, K., Harris, A., & Strauss, T. (2010). *Leading school turnaround: How successful leaders transform low-performing schools*. John Wiley & Sons.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130.

Marks, H. M., & Printy, S. M. (2003). Principal Leadership and School Performance: An Integration of Transformational and Instructional Leadership. *Educational Administration Quarterly*, *39*, 370–397.
Meyers, C. V., & Murphy, J. (2007). Turning around Failing Schools: An Analysis. *Journal of School Leadership*, *17*, 631–659.
Mintrop, H., & MacLellan, A. M. (2002). School Improvement Plans in Elementary and Middle Schools on Probation. *The Elementary School Journal*, *102*, 275–300.
Murphy, J., & Meyers, C. V. (2007). *Turning around failing schools: Leadership lessons from the organizational sciences*. Corwin Press.
New Teacher Center. (2014). *Validity and reliability of the North Carolina Teacher Working Conditions survey.* (No. BRF-VLRL-USNC-1405-EN). Retrieved from https://ncteachingconditions.org/research
Papay, J. (2015, November 12). *The Effects of School Turnaround Strategies in Massachusetts, The Effects of School Turnaround Strategies in Massachusetts*. Presented at the 2015 Fall Conference: The Golden Age of Evidence-Based Policy, 2015 Fall Conference: The Golden Age of Evidence-Based Policy.
Papay, J., & Hannon, M. (2018, November 8). *The Effects of School Turnaround Strategies in Massachusetts*. Presented at the 2018 APPAM Fall Research Conference: *Evidence for Action: Encouraging Innovation and Improvement*, Washington, D.C.
Perry, B. L., & Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, *79*, 1067–1087.
Peurach, D., & Neumerski, C. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*, 379–420.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*, 417–458.
Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, *94*, 247–252.
Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher Collaboration in Instructional Teams and Student Achievement. *American Educational Research Journal*, *52*, 475–514.
Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, *50*, 4–36.
Sansone, D. (2018). Beyond Early Warning Indicators: High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics*, *0*. https://doi.org/10.1111/obes.12277
Schueler, B. E., Goodman, J. S., & Deming, D. J. (2017). Can States Take Over and Turn Around School Districts? Evidence From Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*, *39*, 311–332.
Schweig, J. (2014). Cross-Level Measurement Invariance in School and Classroom Environment Surveys: Implications for Policy and Practice. *Educational Evaluation and Policy Analysis*, *36*, 259–280.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. *The Annual Meeting of the Psychometric Society. Iowa City, IA. 1980*.

Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, S., & Weinstein, T. (2016). The Impact of Turnaround Reform on Student Outcomes: Evidence and Insights from the Los Angeles Unified School District. *Education Finance and Policy*, *11*, 251–282.

Stuit, D. A. (2010). *Are Bad Schools Immortal? The Scarcity of Turnarounds and Shutdowns in Both Charter and District Sectors*. Thomas B.

Sun, M., Penner, E. K., & Loeb, S. (2017). Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants. *American Educational Research Journal*, *54*, 607–643.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A Review of School Climate Research. *Review of Educational Research*, *83*, 357–385.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Viano, S., Pham, L. D., Henry, G. T., Kho, A., & Zimmer, R. (2020). What Teachers Want: School Factors Predicting Teachers' Decisions to Work in Low-Performing Schools. *American Educational Research Journal*.

Zimmer, R., Henry, G. T., & Kho, A. (2017). The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones. *Educational Evaluation and Policy Analysis*, *39*, 670–696.

Appendix

Table A-2-1. Random partition crossvalidation summary statistics

	RMSE	SD	t	Recall	SD	t	Accuracy	SD	t
10^1	13.572	0.0634	214.0694	0.2392	0.0417	5.7403	0.9449	0.0031	307.7850
10^1.1	13.572	0.0601	225.8236	0.2420	0.0429	5.6358	0.9448	0.0031	302.8205
10^1.2	13.571	0.0602	225.4319	0.2436	0.0429	5.6836	0.9447	0.0031	303.7621
10^1.3	13.574	0.0606	223.9934	0.2406	0.0432	5.5733	0.9448	0.0031	303.7942
10^1.4	13.572	0.0583	232.7959	0.2397	0.0417	5.7482	0.9448	0.0031	306.7532
10^1.5	13.571	0.0598	226.9398	0.2410	0.0422	5.7068	0.9448	0.0031	306.7532
10^1.6	13.574	0.0589	230.4584	0.2441	0.0436	5.6012	0.9446	0.0031	301.7891
10^1.7	13.569	0.0610	222.4426	0.2419	0.0428	5.6492	0.9449	0.0031	300.9236
10^1.8	13.575	0.0595	228.1513	0.2401	0.0425	5.6521	0.9447	0.0031	302.7885
10^1.9	13.574	0.0623	217.8812	0.2397	0.0421	5.6909	0.9447	0.0032	298.9557
10^2	13.576	0.061	222.193	0.239	0.042	5.654	0.9448	0.0031	300.892
10^2.1	13.575	0.061	223.641	0.238	0.041	5.738	0.9448	0.0032	294.330
10^2.2	13.581	0.061	221.912	0.237	0.042	5.704	0.9447	0.0033	290.677
10^2.3	13.580	0.061	221.895	0.238	0.042	5.625	0.9446	0.0032	296.113
10^2.4	13.585	0.063	216.667	0.239	0.043	5.530	0.9444	0.0032	295.125
10^2.5	13.585	0.060	226.795	0.235	0.041	5.697	0.9446	0.0032	296.113
10^2.6	13.588	0.060	227.224	0.236	0.041	5.748	0.9442	0.0032	295.987
10^2.7	13.591	0.061	222.439	0.237	0.041	5.734	0.9440	0.0032	298.734
10^2.8	13.595	0.059	230.034	0.236	0.041	5.814	0.9438	0.0032	294.019
10^2.9	13.595	0.059	230.034	0.242	0.042	5.820	0.9432	0.0031	308.235
10^3	13.599	0.058	235.277	0.246	0.042	5.926	0.9428	0.0032	293.707
10^3.1	13.613	0.059	230.729	0.254	0.041	6.160	0.9419	0.0029	323.677
10^3.2	13.628	0.057	239.930	0.262	0.039	6.642	0.9411	0.0026	366.187
10^3.3	13.646	0.059	230.507	0.269	0.036	7.497	0.9406	0.0023	410.742
10^3.4	13.670	0.061	225.578	0.270	0.036	7.499	0.9405	0.0022	435.417
10^3.5	13.698	0.057	240.738	0.277	0.032	8.783	0.9401	0.0020	477.208
10^3.6	13.732	0.061	223.648	0.319	0.023	13.664	0.9428	0.0016	592.956
10^3.7	13.765	0.058	235.702	0.331	0.022	14.863	0.9420	0.0015	623.841
10^3.8	13.806	0.059	234.397	0.347	0.025	13.782	0.9411	0.0016	580.926
10^3.9	13.870	0.059	234.687	0.362	0.019	19.261	0.9403	0.0013	706.992
10^4	13.952	0.058	241.802	0.364	0.014	25.447	0.9397	0.0011	854.273
10^4.1	14.063	0.059	239.574	0.361	0.011	32.761	0.9390	0.0011	885.849
10^4.2	14.186	0.057	247.143	0.359	0.010	36.353	0.9388	0.0010	948.283
10^4.3	14.343	0.060	238.652	0.360	0.010	36.755	0.9390	0.0010	939.000
10^4.4	14.580	0.058	251.379	0.360	0.010	36.731	0.9390	0.0010	978.125
10^4.5	14.917	0.059	251.551	0.349	0.010	35.222	0.9383	0.0010	938.300
10^4.6	15.355	0.059	262.031	0.344	0.011	31.852	0.9377	0.0011	884.623

10^4.7	16.010	0.058	276.034	0.334	0.011	31.600	0.9365	0.0011	883.491
10^4.8	16.939	0.057	295.105	0.314	0.014	22.121	0.9319	0.0013	745.520
10^4.9	18.036	0.064	283.140	0.237	0.009	26.637	0.9246	0.0010	924.600
10^5	19.195	0.063	302.760	0.237	0.009	27.462	0.9246	0.0010	943.469

Means and standard deviations of RMSE, recall, and accuracy rates over 1,000 bootstrapped samples (80% of full sample of schools) for each value of λ. Recall rate is the primary fit statistic of interest and represents the proportion of the bottom 5 percent accurately predicted by the model. Accuracy rate is the proportion of the total population accurately predicted by the model. RMSE represents the unexplained variance in the continuous school performance variable. The test statistics listed are the mean of each measure divided by the standard deviation. The goal is to maximize recall and accuracy while minimizing RMSE and maximizing each test statistic.

Table A-2-2. Rotated factor loadings (orthogonal rotation)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	ψ
Enrollment (rev)	0.023	-0.136	0.558	0.475	-0.150	0.422
Minority students	0.969	0.206	0.133	-0.004	0.044	0.000
Black students	0.793	0.200	0.106	0.045	0.246	0.257
Econ. dis. in-migration	0.322	0.161	0.295	0.666	-0.080	0.332
Disabled in-migration	-0.154	-0.040	0.277	0.245	-0.005	0.838
Student-teacher ratio	-0.121	-0.094	-0.809	-0.197	0.038	0.281
Novice teachers	0.402	0.258	0.003	0.035	0.024	0.770
Alternative entry teachers	0.233	0.321	-0.133	-0.031	0.579	0.489
Teacher supplement (rev)	-0.246	0.066	0.033	0.656	0.213	0.459
Short-term suspension rate	0.282	0.328	0.003	0.077	0.644	0.392
Per pupil expenditures	0.186	0.013	0.812	-0.016	0.023	0.304
Community engagement (rev)	0.244	0.951	0.084	0.131	0.111	0.000
Student conduct (rev)	0.155	0.728	-0.087	-0.154	0.182	0.382

Factor loadings with orthogonal varimax rotation

Chapter 3

Are There Turnaround Principals?

Turning around low-performing schools may require a different set of leadership skills and experiences than leading better performing schools. Qualitative research suggests the existence of “turnaround principals” — leaders with the necessary talent for improving low-performing schools (Meyers & Hambrick Hitt, 2017; Peck & Reitzug, 2014). Success leading a higher performing school would not necessarily translate into success in a turnaround school because turnaround requires a set of competencies specifically conducive to the challenges of turnaround (Steiner & Hassel, 2011). However, school districts hiring principals for their low-performing schools have little information about whether a given candidate has the requisite skills and talents to be a turnaround principal. Instead, available information is often limited to prospective principals’ education, certification, and past experience.

Meanwhile, principals in high-needs and high-minority schools are also more likely to turn over than their counterparts in other schools (Battle, 2010; Béteille et al., 2012; Fuller & Young, 2009; Gates et al., 2006; Loeb et al., 2010), highlighting the particular importance for high-needs schools to hire principals who will be effective early in their tenure. Though a wealth of research has demonstrated the differences in principal preparation and retention between schools with high levels of poverty and minority populations and schools with lower levels of student poverty and minority rates, none of this research has focused on differences in turnaround schools in particular. As the complexity of the principalship has increased with the expansion of federal and state accountability (Kafka, 2009; Tekleselassie & Villarreal III, 2011), the incentives and deterrents to taking the reins of a turnaround school in particular may be more even complex than entering other high-needs schools. And in fact, there is evidence that principals facing higher accountability pressure are more likely to turn over than principals without those pressures (DeAngelis & White, 2011).

Existing research has attempted to identify the competencies associated with successful turnaround principals. These studies have focused on the attributes and decisions of principals leading successful turnarounds or reviews of case studies of low-performing schools undergoing turnaround (Copeland & Neeley, 2013; Meyers & Hambrick Hitt, 2017). These studies and other policy reports related to school turnaround have highlighted that turnaround is a distinct discipline requiring a school leader with a correspondingly distinct set of skills (Calkins, Guenther, Belfiore, & Lash, 2007; Copeland & Neeley, 2013; Duke, 2004; Duke & Salmonowicz, 2010; Meyers & Hambrick Hitt, 2017). No research to date has empirically examined whether principals with successful track records in low-performing schools are more effective than principals who have successfully guided non-low-performing schools. Given the limited information available to districts hiring principals for turnaround schools, knowing whether past success in a low-performing school is predictive of future success would allow district officials to draw on available information in narrowing applicant pools and making hiring decisions in their lowest performing schools. The first three questions in this chapter are descriptive questions intended to document the frequency of turnaround principals and their mobility patterns. The fourth question attempts to move toward a causal estimate of the effect of turnaround principals in low-performing schools. Specifically, I ask:

(1) How frequently are principals successful in low-performing schools and how often do they sustain that success over multiple years?

(2) Do successful principals of low-performing schools leave at higher rates than successful principals of non-low-performing schools?

(3) Where do principals with demonstrated prior success in low-performing schools ("turnaround principals") go after leaving the formerly low performing school?

(4) Are these turnaround principals more effective at leading low-performing schools than principals who have successfully led non-low-performing schools?

Literature Review

In this section, I overview three areas of literature related to turnaround principals. I begin by describing relevant literature on school turnaround in order to characterize successful turnaround. I then review the literature on principal effectiveness to set the foundation for how principals contribute to school performance. Finally, I describe the literature on turnaround principals on which this study will build.

School turnaround

Two key principles underlie school turnaround. The first is that improvement under turnaround is rapid and dramatic, while the second is that turnaround should be sustainable—after achieving rapid improvements, schools need to sustain those improvements into future years (Public Impact, 2009). The goal for change to be rapid and dramatic distinguishes turnaround from previous generations of school reform models, which assumed an incremental approach that would occur over three to five years (Herman et al., 2008; Peurach & Neumerski, 2015). The literature typically defines rapid change as making dramatic improvements within two or three years (Aladjem et al., 2010; Herman et al., 2008; Rhim & Redding, 2014). The goal for sustainable change is not distinct from goals under prior reform models, but it has proven to be an ambitious objective as schools have struggled to sustain improvements three years after turnaround (Duke & Landahl, 2011; Hochbein, 2012).

While the two-year timeline for change constitutes an established goal for the "rapid" component of school turnaround, there are no settled definitions of what makes a turnaround successful or sustainable (Trujillo & Renee, 2012). Existing studies have employed varied definitions of success, including moving above the median school in the state (Stuit, 2010), making one-year gains in the top 50 percent of gainers, and making consistent annual gains in both reading and math (Aladjem et al., 2010). While many studies discuss sustainability as a relevant criterion for turnaround, they do not offer specific barometers for what constitutes a sustainable turnaround (Murphy & Bleiberg, 2019).

While there is no agreed-upon threshold for success, a handful of turnarounds under RttT and SIG have yielded positive effects on student achievement that were sustained over multiple years. For example, Massachusetts' SIG program produced significant positive effects in math and ELA achievement for students in 35 low-performing schools. These effects emerged in the first year of turnaround and grew through the fourth year (Papay & Hannon, 2018). Low-performing schools in Tennessee, where the state undertook two parallel turnaround initiatives under RttT, improved under one turnaround model but stagnated under the other (Zimmer et al., 2017). The significant increases in math, reading, and science achievement under the successful Innovation Zone (iZone) district-within-a-district model persisted for six years (Pham et al., 2018). In Ohio, Carlson & Lavertu (2018) found some evidence that positive effects of SIG on reading scores may have persisted for four years, though these estimates were not consistently significant across specifications, and there were no significant effects on Year 4 math scores.

Not all studies of turnaround since RttT and SIG have found positive effects—even in the first years of reform. A large-scale evaluation of SIG grants in 22 states found null effects (Dragoset et al., 2017) and a study of Texas SIG schools found negative effects (Dickey-Griffith, 2013). North Carolina's RttT turnaround initiative produced mixed effects, with positive effects in high schools and the very lowest performing schools, and some negative effects in elementary and middle schools (Heissel & Ladd, 2018; Henry et al., 2015). In a subsequent statewide turnaround initiative in North Carolina, student achievement growth in treated schools declined and teacher turnover increased in the second year of reform (Henry & Harbatkin, 2018).

The heterogeneity of effects across interventions may stem in part from school leadership. There is broad agreement in the literature that leadership plays a critical role in school improvement (Aladjem et al., 2010; Calkins et al., 2007; Dodman, 2014). Qualitative research shows that traditional public schools that have overcome challenges around low performance and high poverty benefit from the leadership of strong principals with a clear and articulated vision for success (Calkins et al., 2007), and who employ distributed leadership to carry out the complex tasks associated with turnaround (Elmore, 2000; Leithwood et al., 2009).

Principal effectiveness

A large literature demonstrates that school leadership matters for school performance (Hallinger & Heck, 1998; Witziers et al., 2003). Principals exert a small direct effect on student achievement (Witziers et al., 2003), though much of their influence is likely to occur through mediating factors such as teacher retention, school climate, goal-setting and expectations, establishing and fostering collaborative processes, and organizational decisions (Finnigan & Stewart, 2009; Hallinger & Heck, 1998; Jacobson et al., 2005; Leithwood & Jantzi, 1990). These mediating factors are particularly relevant in low-performing schools, where successful turnaround may demand sweeping changes to the teaching workforce, a strong vision for change, and adoption of purposive practices intended to improve school climate (Aladjem et al., 2010).

Effective school leadership requires strong organizational and time management skills (Grissom, Loeb, et al., 2015; Grissom & Loeb, 2011), and successful leaders may prioritize their time to focus on activities related to organizational and time management. For example, a study of principal time use found that schools in which principals prioritized organizational management activities experienced larger student achievement gains, scored higher on climate ratings from staff, and were perceived as safer by parents (Horng et al., 2010). Employing effective organizational and time management may be particularly important undertakings in low-performing schools, which struggle with instabilities related to teacher mobility—especially among effective teachers—student transfers, and high rates of novice teachers and leaders (Boyd et al., 2005; Clotfelter et al., 2005; Goldhaber et al., 2015; Guarino et al., 2006; Hanushek et al., 2004; Mao et al., 1997; McGee, 2004).

Although school leadership is critical to student and school outcomes, principal effectiveness is highly variable (Branch et al., 2009, 2012; Coelli & Green, 2012; Dhuey & Smith, 2018; Grissom, Kalogrides, et al., 2015). Principal effectiveness is most variable at high-poverty schools (Branch et al., 2009). The critical decision of selecting a principal for a turnaround school therefore has exceptionally high stakes in a labor market with both highly effective and highly ineffective principals. Identifying an appropriate turnaround principal may be especially important because school leadership plays a key role in influencing some of the school-level factors with which low-performing schools struggle most known, such as teacher satisfaction and retention—and evidence shows that principal effectiveness is associated with higher teacher satisfaction and retention, especially among effective teachers (Grissom, 2012;

Grissom & Bartanen, 2019). These positive impacts of principal effectiveness are larger in disadvantaged schools (Grissom, 2011). Principal effectiveness is therefore particularly important in low-performing schools, which may be more responsive to more or less effective principals.

Turnaround principals

Qualitative research on turnaround principals and leadership in low-performing schools suggests a given principal's effectiveness may vary by school context. Principals who successfully lead higher performing schools may not be as effective in low-performing schools, which require leaders to possess a distinct set of skills and traits (Copeland & Neeley, 2013; Duke, 2004; Duke & Salmonowicz, 2010; Finnigan & Stewart, 2009; Harris, 2002; Jacobson et al., 2005; Meyers & Hambrick Hitt, 2017). Drawing from 18 studies of turnaround principals between 2000 and 2015, Meyers & Hambrick Hitt (2017) identified 12 domains of successful turnaround leadership. Each of these domains fit into one of three broad functions of the turnaround principal: (1) vision and strategic leadership, (2) building capacity with support and accountability, and (3) shaping school culture (Meyers & Hambrick Hitt, 2017).

While the particular traits deemed necessary for a successful turnaround vary somewhat across studies, skills pointing toward effectiveness in these three areas are consistent across studies and contexts. For example, setting clear vision and goals emerged as a vital mission for turnaround principals across school levels and settings. In a case study of 10 low-performing elementary schools in Chicago, Finnigan & Stewart (2009) found that effective principals articulated high expectations to staff through a regular and consistent focus on collective improvement goals. A study of 10 UK secondary schools that were low-performing or high-poverty found successful school leaders ensured schoolwide alignment to a shared set of values, often through symbolic gestures and actions to set the course for improvement and help staff to remain focused on school goals (Harris, 2002). A study of seven schools in New York state spanning multiple school levels and urbanicities found that effective principals were skilled at leveraging external accountability pressures to bring clarity to direction-setting and school goals (Jacobson et al., 2005). Because school turnaround is intended as a rapid and dramatic process (Herman et al., 2008), school goals should be ambitious, and successful turnaround leaders poised to take responsibility for meeting these challenging goals (Public Impact, 2008). Turnaround principals must therefore be adept at establishing a sense of urgency to achieve school goals (Dodman, 2014; Meyers & Hambrick Hitt, 2017).

The second broad pillar of successful turnaround—building capacity with support and accountability—comprises investing in staff development, developing systems and structures to facilitate development, and holding staff accountable. Capacity building involves strategic investment in resources aligned with school goals, developing school structures to facilitate collaboration toward common goals, supporting teachers in delivering data-driven instruction, hiring effective teachers and leaders, and supporting and developing staff. In the study of Chicago elementary schools, successful principals targeted school resources toward an explicit school vision and provided teachers with supports intended to guide their efforts in achieving those goals (Finnigan & Stewart, 2009). Similarly, successful principals in UK secondary schools invested in staff development through in-service training, visits to other schools, and facilitating peer supports (Harris, 2002). Distributed leadership was a common thread throughout each of these and other successful turnaround interventions. By empowering individuals as leaders, turnaround principals fostered a supportive environment for teachers to build their individual capacities while simultaneously shoring up school systems and structures to facilitate continued growth. For example, successful principals in Chicago employed controlled distributed leadership by empowering assistant principals and teacher leaders to support other teachers in examining student data for data-driven instruction (Finnigan & Stewart, 2009). Effective principals in New York worked to develop a set of interconnected committees charged with decision making intended to improve student learning. They also instituted structures to facilitate collaboration and teamwork, such as common planning time (Jacobson et al., 2005). Together, these strategies highlight the intersecting decisions principals make in their efforts to guide turnaround. Decisions around capacity building cannot be independent from vision-setting. Successful capacity building also requires follow-through by committed assistant principals and teacher leaders who buy into school vision and goals.

Finally, case studies suggest that successful turnaround requires meaningful change to school culture. In Chicago, successful turnaround principals promoted positive school culture by creating opportunities for regular and meaningful collaboration, improving physical structures that hinder school climate, and fostering relational trust (Finnigan & Stewart, 2009). Similarly, a study of school culture in 12 improving schools in Canada found that transformational principals were able to alter organizational structures to facilitate positive school culture and relational trust (Leithwood & Jantzi, 1990). A study of turnaround schools in California found that principals worked deliberately to change the culture of teacher collaboration through improved professional learning communities, scheduling to facilitate collaboration, and promoting common instructional practices (Huberman et al., 2011). Meyers & Hambrick Hitt (2017) argued that successfully transforming deep-seated conventions and structures that shape school culture may require a leader with a particular set of skills and attributes such as a willingness to disrupt complacency in the face of resistance to change.

The multidimensional nature of school turnaround points to the need for a leader who has the requisite skills to effectively carry out each of these objectives. Districts seeking to fill principalships in low-performing schools are tasked with identifying individuals with the particular set of skills necessary for articulating and promoting a shared vision for the school, building staff capacity through individual and structural supports and resources, and transforming school culture. But at the time of hiring, districts have information only on qualifications such as education, certification, and years of experience, as well as data on performance in prior schools. This study draws on the latter information to make inferences about which prospective principals might possess these skills as "turnaround principals."

Methods

Data and sample

To examine whether there are "turnaround principals" who are more effective leaders in low-performing schools than other principals, I draw from 13 years of statewide longitudinal data from North Carolina from 2005-06 to 2017-18. The administrative database contains longitudinal data on all public school students, classes, and personnel in the state. I supplement the administrative data with the Teacher Working Conditions (TWC) survey, administered biennially in even years during the outcome period. The analytic sample includes principals, teachers, and students in more than 2,000 regular public schools for which I can calculate principal value-added from 2009-10 through 2017-18. I use the data from four years prior to the 2009-10 school year (2005-06 through 2008-09) to identify the first year a principal is in a school. I exclude charters, alternative schools, special education schools, and hospital schools, where turnaround may require a different set of skills than regular public schools. In sum, from

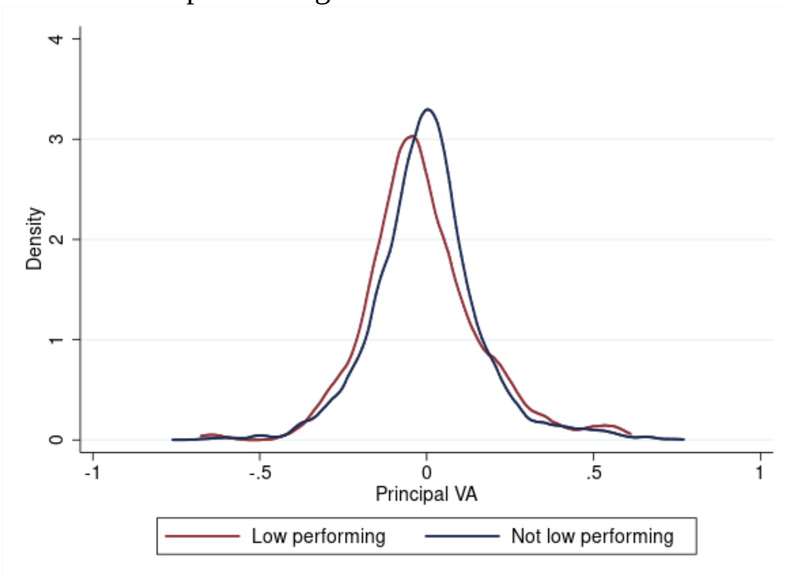
2009-10 through 2017-18, I have data from about 4,800 unique principals, 2 million students, and 177,000 teachers.

Measures

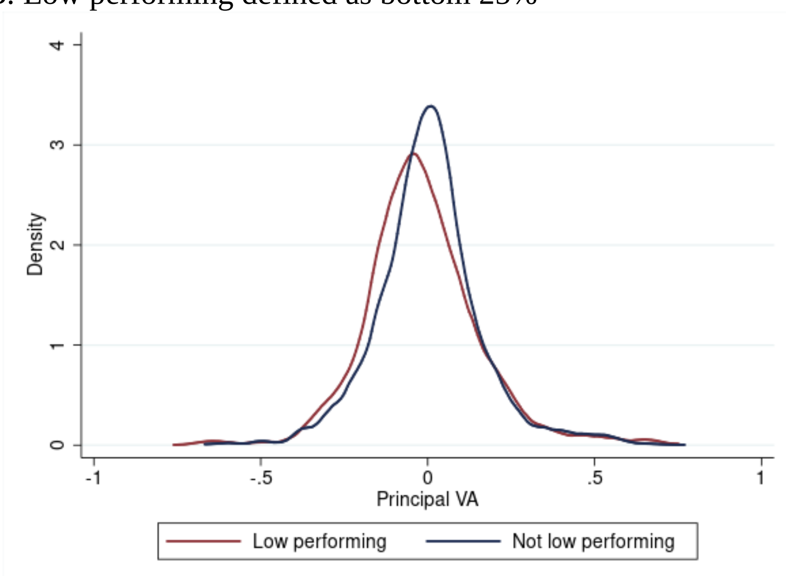
Principal value-added. Before I can classify a principal as successful or unsuccessful in order to tag turnaround principals, I need to define a measure of principal effectiveness in the schools from which principals are transferring (sending schools). To create this measure, I draw from math and reading scores, standardized by subject, grade, and year. Specifically, I estimate the drift-adjusted principal value-added measure described by Bartanen (2020), which modifies the teacher value-added estimator developed by Chetty et al., (2014) to include school and principal fixed effects. The leave-year-out approach allows principal value-added to change over time, a necessary condition for identifying turnaround principals under my six definitions, as I describe later. There are three steps to the process. The first step is to residualize student test scores by running an ordinary least squares regression predicting student test score as a function of lagged test score, a vector of student characteristics, a vector of school characteristics, and principal and school fixed effects. These residuals provide information about student test score growth not explained by observables and stable principal quality. The second step is to extract a variable equal to the student residuals plus the principal fixed effect and collapse it to the principal-by-year level. This variable represents the school mean year-to-year student growth, controlling for observables. The third step is to predict principal value-added in year t as a function of the principal-by-year residuals from all other years (Bartanen, 2020; Chetty et al., 2014). The jackknife approach avoids bias that would be introduced by including the same estimation error on both sides of the equation (Chetty et al., 2014; Jacob et al., 2010). School-level covariates included in the first step include minority percentage, per pupil expenditures (PPE) and PPE squared, and enrollment and enrollment squared. Student-level covariates include indicators for student race, female, over-age for grade, disability, English learner, and economically disadvantaged. Figure 3-1 shows the distribution of principal effects by a school's low-performing status in the year prior to the first year of the principal's tenure. There is considerable overlap between principal value-added in schools that were low performing in the year prior to a principal's entry and schools that were not. While proficiency rates require a student to cross the proficiency threshold regardless of her baseline achievement, the value-added measure captures student growth. While a school value-added measure would also capture student growth, the principal measure also partials out any between-school variation that is associated with principal effectiveness and student achievement (Bartanen, 2020). Additionally, there is evidence that school and principal value added are not strongly correlated with one another (Chiang et al., 2016).

Figure 3-1. Distribution of principal value-added by school low-performing status

Panel A. Low performing defined as bottom 10%



Panel B. Low performing defined as bottom 25%



Of particular importance to interpreting these estimates is that the two-way fixed effects model estimates each principal's deviation from the mean effect of a network of principals connected by principal movement across schools (Burkhauser, 2017; Chiang et al., 2016). These connected networks are groups of principals and schools connected by a principal's transfer from one school to another in the network. Principals who remained in a single school for the entirety of the study period are disconnected from all other principals and I cannot estimate value-added for them. I drop 232 principals who are not in connected networks from the value-added estimation. While I can estimate value-added for any principal in a school with at least two principals, many connected networks are quite small. For example, 21 percent of principals are in networks containing just one school. The value-added estimate for these principals is therefore estimated relative only to other principals in their school—a limitation of this estimator. Table 3-1 shows the distribution of connected network sizes.

Table 3-1. Connected network sizes

Network size (N schools)	N networks of this size	Mean number of principals per network	Total number of principals	Percent of principals
1	360	2.5	844	21%
2	112	4.0	414	10%
3	44	6.0	254	6%
4	26	8.0	199	5%
5	5	10.9	54	1%
6	8	12.6	98	2%
7	5	13.7	67	2%
8	3	16.1	47	1%
9	1	19.0	19	0%
10	3	19.9	58	1%
12	3	24.2	72	2%
13	3	22.7	68	2%
14	1	27.0	27	1%
15	1	22.0	22	1%
16	2	35.5	71	2%
20	1	36.0	36	1%
21	1	45.0	45	1%
22	1	38.0	38	1%
25	1	46.0	46	1%
28	1	50.0	50	1%
36	1	73.0	73	2%
37	1	84.0	84	2%
84	1	170.0	170	4%
87	1	165.0	165	4%
100	1	222.0	222	5%
402	1	825.0	825	20%
Any	588	6.9	4068	100%

I count a principal as effective if her value-added estimate is greater than .25 standard deviations above zero. By setting the effectiveness threshold higher than zero, I aim to exclude principals who are higher than zero by chance. While .25 SDU is still an arbitrary cutoff, it requires that principals are above average in their connected networks in order to count as effective without excluding too many principals to estimate the effect of turnaround principals. Across all principals from 2010 through 2018, 2,066 of 7,763 principals with value-added scores above zero have scores between zero and .25 SDU above zero.

Dependent variables. The primary outcome of interest is math and reading standardized test scores. Students in North Carolina take end-of-grade (EOG) exams in math and reading each year from third through eighth grade, and end-of-course (EOC) exams in Math I and English II. This EOC test structure began in 2013 with adoption of the Common Core State Standards. Prior to 2013, students took math and ELA EOCs in Algebra I, Algebra II, and English I.

In addition to measuring the influence of turnaround principals on student test scores, I also examine more proximal outcomes that may contribute to a change in test scores such as teacher turnover and chronic absenteeism. I operationalize teacher turnover as a binary indicator denoting whether a teacher leaves the school at any point during or at the end of school year *t*, regardless of pathway out. I operationalize student chronic absenteeism as a binary indicator representing whether a student was absent for at least 10 percent of enrolled days.

Turnaround principal. I identify potential turnaround principals in six ways, all based on performance in prior low-performing schools. I exclude principal-school observations in which I do not observe the principal's first year at a given school because I classify a school as low performing based on its performance in the year prior to the principal's first year. I am interested in the effectiveness of principals who entered a low-performing school—not the effectiveness of principals who oversaw a school's drop into low performance.

Underlying the definitions of turnaround principals are the two key tenets of school turnaround—first, the extent to which turnaround is rapid and dramatic, and second, the extent to which it is sustainable. I operationalize "rapid and dramatic" as achieving a score of more than .25 standard deviations above 0 on my principal value-added measure in a prior school. I measure sustainability by the number of years in which the principal reached the growth threshold in a previously low-performing school. The six definitions of turnaround principals are situated on a continuum ranging from the most liberal definition to the most conservative on these two tenets.

Across all definitions, the school achievement threshold I use to categorize a principal as effective or not is a principal value-added score at least .25 standard deviations greater than zero. About one-third of principal-by-year observations had value-added estimates meeting this criterion, placing them in the "effective" category for a given year. These numbers vary little by the school's low-performing status, as shown in Table 3-2.

Table 3-2. Frequency of effective principal by low-performing status in principal's first year in a school, 2010-2018

	Effective	Ineffective	Obs
Low-performing defined as bottom 10%			
Low-performing	33.7%	66.3%	424
Non-low-performing	34.3%	65.7%	3209
Low-performing defined as bottom 25%			
Low-performing	32.4%	67.6%	965
Non-low-performing	34.9%	65.1%	2668

NOTE: Principal-by-school-by-year observations. School status is based on school performance in the year immediately prior to a principal's first year in the school.

The most liberal definition is a principal who were effective for at least one year at any time in a school that was low performing the year before she began the principalship. This definition will identify the largest possible sample of turnaround principals by capturing any principal who was successful for at least one year—either immediately or later—in a low-performing school, regardless of tenure length at that school. To hone in on a narrower conceptualization of "rapid," the next definition for a turnaround principal is one that is effective for at least one year within two years of entering the school. This timeline aligns with the school turnaround literature that calls for turnaround to occur within the first two years (Aladjem et al., 2010). About 41 percent of principals in low-performing schools (under both the bottom 10% and bottom 25% low-performing designations) were effective in their first two years, compared with 44 to 45 percent in the non-low-performing schools, as I show in Table 3-3.

Table 3-3. Frequency of effective principal within first two years in a school by low-performing status in principal's first year in a school, 2010-2018

	Effective in first 2 years	Not effective in first 2 years	Obs
Low-performing defined as bottom 10%			
Low-performing	40.8%	59.2%	338
Non-low-performing	43.9%	56.1%	2713
Low-performing defined as bottom 25%			
Low-performing	40.5%	59.5%	783
Non-low-performing	44.7%	55.3%	2268

NOTE: Principal-by-school observations for first two years in school. School status is based on school performance in the year immediately prior to a principal's first year in the school. Sample restricted to principals that were in observed school for at least two years.

In each subsequent definition of turnaround principal, I add one additional year of principal effectiveness (i.e., at least 2 years, at least 3 years) and alternate between requiring that principals reach this threshold at any point during their tenure in a given school, and that they effective at least once within the first two years in a school. The most conservative definition requires that principals meet the value-added effectiveness threshold for at least three years in a low-performing school and that the first year of effectiveness occurred within their first two years of entering the school. Fewer principals meet the criteria as the definitions move down the continuum in Figure 3-2.

Figure 3-2. Definitions of turnaround principal

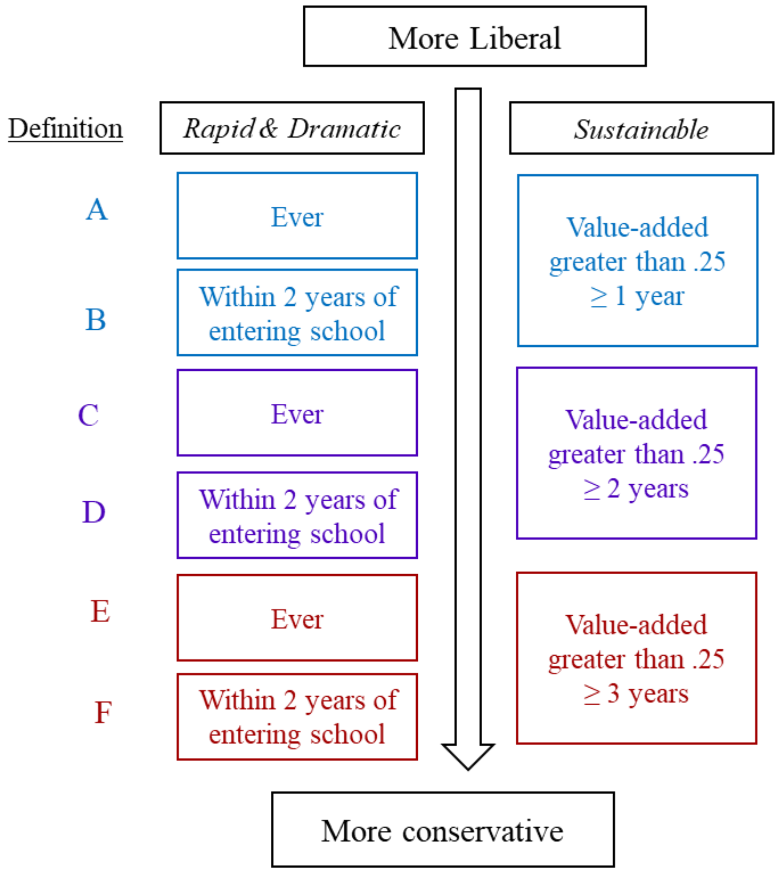


Table 3-4 shows the proportion of principals in low-performing and non-low-performing schools who are effective for at least one, two, and three years in that school. Less than 40 percent of principals in low-performing schools are effective for at least one year of their tenure in that school, while 43-44 percent of principals of non-low-performing schools are effective for at least one year. Within both types of schools, fewer principals are effective for at least three years—less than a fourth of principals who spent at least three years in a low-performing school were effective for that many years, and about a fourth of principals who spent at least three years non-low-performing schools were effective for at least three years.

Table 3-4. Frequency of effective principal for 1-3 years by low-performing status in principal's first year in a school, 2010-2018

A. Low-performing as bottom 10%

	All principals			Principals with at least 3 years		
	1+ year	2+ years	3+ years	1+ year	2+ years	3+ years
Non-LPS	0.434	0.305	0.192	0.446	0.345	0.251
	(0.496)	(0.460)	(0.394)	(0.497)	(0.476)	(0.434)
LPS	0.383	0.250	0.139	0.401	0.301	0.216
	(0.487)	(0.433)	(0.347)	(0.491)	(0.460)	(0.412)
Observations	4620	4620	4620	3486	3486	3486

B. Low-performing defined as bottom 25%

	All principals			Principals with at least 3 years		
	1+ year	2+ years	3+ years	1+ year	2+ years	3+ years
Non-LPS	0.439	0.311	0.195	0.451	0.349	0.252
	(0.496)	(0.463)	(0.396)	(0.498)	(0.477)	(0.434)
LPS	0.395	0.262	0.161	0.412	0.313	0.235
	(0.489)	(0.440)	(0.368)	(0.492)	(0.464)	(0.424)
Observations	4620	4620	4620	3486	3486	3486

NOTE: Principal-by-school observations. Numbers in cells reflect the proportion of principals who were effective, defined as having a principal VA score of more than .25 SD >0, for at least one, two, and three years. School's low-performing status determined in year prior to observed principal's entry.

Putting together the two sets of requirements to meet the "effective principal" and "turnaround principal" definitions—effective in the first two years and effective for at least one, two, and three years in the school—Table 3-5 shows the number of principals meeting each of those definitions across the full sample of schools. Principals in low-performing schools meet the most liberal definition of effectiveness (being effective for at least one year) at similar rates to those in non-low-performing schools. The number of principals meeting the criteria for effectiveness decreases as the definition becomes more conservative, and a small gap between low-performing and non-low-performing schools emerges. Under the most conservative definition of effectiveness, 26 percent of principals of low-performing schools meet the criteria, while 29 percent of principals of non-low-performing schools meet the criteria.

Table 3-5. Sending school frequency of turnaround and effective principals by definition and *sending* school status in principal's first year, 2010-2018

A. Low-performing defined as bottom 10%

	Low-performing school		Non-low-performing school	
	Total	Prop	Total	Prop
Definition A	151	0.262	1484	0.256
Definition B	135	0.335	1103	0.344
Definition C	107	0.330	972	0.352
Definition D	106	0.327	957	0.347
Definition E	61	0.262	620	0.288
Definition F	61	0.262	616	0.286
N	577		5796	

B. Low-performing defined as bottom 25%

	Low performing		Not low performing	
	Total	Prop	Total	Prop
Definition A	357	0.267	1278	0.254
Definition B	300	0.326	938	0.348
Definition C	248	0.327	831	0.357
Definition D	244	0.321	819	0.352
Definition E	156	0.275	525	0.288
Definition F	154	0.272	523	0.287
N	1337		5036	

NOTE: Principal-by-school observations. School status is based on school performance in the year immediately prior to a principal's first year in the school. Principals in first columns (low-performing school) tagged as turnaround principals at any subsequent schools. Principals in second columns (non-low-performing school) tagged as other effective principals at any subsequent schools. The N for each table represents the N for Definition A. The N's become smaller for each subsequent definition because the denominator includes only principals who had the opportunity to reach the listed definition in a school. For example, a principal who remained in a school for only one year would be counted as part of the Definition A calculation but not the Definition C calculation because she did not have the opportunity to be effective for at least two years.

If a principal fits one or more of these definitions in a low-performing school, I classify her as a turnaround principal by that definition. If she meets one or more of these definitions in a non-low-performing school, I classify her as an “other effective principal” by that definition. I am then interested in the measuring the effect of that principal in the next school she enters.

To measure the effects of turnaround principals and other effective principals on student and teacher outcomes, I therefore need these principals to move to another school. I code the turnaround principal indicator as 1 if the principal in year *t* was classified as a turnaround principal based on her performance in a low-performing school in a year prior to year *t*. The turnaround principal indicator takes a value of 0 in in schools with principals who (a) previously worked in a low-performing school but did not fit the turnaround principal definition in that school, (b) I never observe in a low-performing school, and (c) I do not observe prior to their entry into a school. I code the other effective principal indicator similarly. It takes a value of 1 if the school’s principal in year t was classified as “other effective principal” based on her performance in a prior non-low-performing school, and a value of 0 in schools with principals who (a) previously worked in a non-low-performing school but did not fit the effective principal definition in that school, (b) I never observe in a non-low-performing school, and (c) I do not observe prior to their entry into that school.

This approach adds considerable noise to the measure. However, dropping all of these school-by-year observations would not be feasible because most principals do not have experience in low-performing schools.[22] Given the limited mobility into low-performing schools during my 10-year time period, the number of turnaround principals in each of these definitions who transfer into new schools is very low, with the most liberal definition including 41 school-

[22] I do try to reduce the amount of noise in the more conservative definitions of turnaround principal by dropping schools with principals who were previously in low-performing schools but did not stay for a long enough period to meet the definition. For example, I code a principal who left after two years in a low-performing school as missing for Definitions E and F because they did not have the opportunity to be effective for at least three years in the school. Including or excluding these principals does not meaningfully change estimates.

by-year observations when I define low performing as the bottom 10 percent of schools and 187 school-by-year-observations when I define low performing as the bottom quartile of schools (Table 3-6). Given the very small number of turnaround principals by the more conservative definitions who move into new schools during the study period, I focus on Definition B in my analysis of turnaround principal effects. I draw from this definition for two reasons. First, it classifies a larger number of principals as turnaround and effective principals than some of the more conservative definitions. Second, unlike Definition A—which identifies more principals—it requires that a principal meet the effectiveness threshold in her first two years in a school. Table 3-5 above underscores the importance of meeting the effectiveness threshold within the first two years in a school. The similar number of principals fitting definitions C and D, and E and F shows that almost all principals who eventually meet these more conservative criteria (i.e., effective for at least two years and effective for at least three years, respectively) reach the effectiveness threshold at least once in their first two years in a school. To that end, rapid effectiveness appears to be particularly important to eventual effectiveness. I therefore want to measure the effects of principals who were not just effective in prior schools—but who exhibited effectiveness early.

Table 3-6. Receiving school frequency of turnaround principals and other effective principals in a low-performing school by definition, 2011-2018

A. Low-performing defined as bottom 10%

	Turnaround principals		Other effective principals	
	Total	Prop	Total	Prop
Definition A	41	0.026	229	0.144
Definition B	30	0.021	160	0.113
Definition C	22	0.016	88	0.065
Definition D	22	0.016	88	0.065
Definition E	12	0.010	47	0.039
Definition F	12	0.010	47	0.039
N	1593		1593	

B. Low-performing defined as bottom 25%

	Turnaround principals		Other effective principals	
	Total	Prop	Total	Prop
Definition A	187	0.049	352	0.092
Definition B	143	0.042	228	0.067
Definition C	96	0.030	129	0.040
Definition D	96	0.030	127	0.040
Definition E	56	0.019	63	0.022
Definition F	56	0.019	63	0.022
N	3814		3814	

NOTE: School-by-year observations for low-performing schools only. School status is based on school performance in the year immediately prior to a principal's first year in the school. Schools in first columns (turnaround principals) had principals who met relevant definition in a prior low-performing school. Schools in second columns (other effective principals) had principals who met relevant definition in non-low-performing school.

Table 3-6 also shows the number of school-by-year observations with principals who previously met each of these definitions in non-low-performing schools (other effective principals). As I describe in the empirical strategy section below, this second set of schools comprises the comparison of interest.

Pathway out. I define four pathways out of the principalship of a given school. *Leavers* are those who leave the North Carolina education system altogether; *demotions* are those who move down to positions such as assistant principal, teacher coach, or teacher (though I cannot observe whether principals are truly demoted or if they choose to move to a new role); *promotions* are those that move to a position generally considered a career advancement for a principal, such as district supervisor, assistant superintendent, superintendent; and *movers* are those who move to the principalship in another school. I then more closely examine the receiving school of movers according to the school's performance level: the bottom 5 percent, 5th-10th percentile, 10th-25th percentile, 25th-50th percentile, 50th-75th percentile, and top 25%. I also separately indicate intra- and inter-district moves.

Other independent variables. I classify schools as low performing based on the state-calculated overall performance composite, which represents the percentage of EOG and EOC exams passed with a score of proficient or above. While proficiency rates do not fully capture a school's performance in a given year and the Every Student Succeeds Act (ESSA, 2015) moved states toward a more nuanced, multidimensional measure of school performance, North Carolina largely used performance composite to identify its lowest performing schools during the study years.[23] I create two definitions of low performing—being in the bottom 10 percent of schools on performance composite during the school year and being in the bottom quartile of schools on performance composite during the school year. Ideally, I would count only the bottom 5 percent of low-performing schools to align with federal requirements for identifying low-performing schools, but there is insufficient movement of effective principals into the bottom 5 percent of schools to measure the effect of turnaround principals during the study period. There are three primary benefits to using performance composite to identify low-performing schools. First, it provides a straightforward definition of low performance that consistently applies across study years. Second, it aligns closely with the state definition of low performance, which varied

[23] I do try to reduce the amount of noise in the more conservative definitions of turnaround principal by dropping schools with principals who were previously in low-performing schools but did not stay for a long enough period to meet the definition. For example, I code a principal who left after two years in a low-performing school as missing for Definitions E and F because they did not have the opportunity to be effective for at least three years in the school. Including or excluding these principals does not meaningfully change estimates.

somewhat across study years but always identified the lowest performing schools based on performance composite—albeit some years with restrictions or additions (see footnote 23). Finally, performance composite was a number readily available to district officials, who would have considered schools with low performance composites as low-performing schools and would have therefore assigned principals accordingly.

In some models, I include a set of school-, principal-, teacher-, and student-level covariates. School-level covariates include per pupil expenditures, average daily membership, student minority percentage, and economically disadvantaged percentage. Principal covariates include gender, race, education level, and years of principal experience. Teacher covariates include gender, race, education level, and teaching experience. Student covariates include gender, race, whether the student is economically disadvantaged, lagged test score, and an indicator for if the student transferred into the school outside of the typical feeder pattern.

Balance checks include principal ratings as measured by the North Carolina Educator Effectiveness System (NCEES), which evaluates principals on seven standards: strategic leadership, instructional leadership, cultural leadership, human resource leadership, managerial leadership, external development (i.e., parent and community involvement) leadership, and micropolitical leadership. Principals receive scores of 1 through 5 on each of these standards, where a 1 indicates the principal did not demonstrate competence in the standard and a 5 indicates the principal's performance exceeded the standard. A score of 3 or above denotes represents proficiency. I include the score on each of the seven dimensions as well as the median score across the dimensions.

Empirical strategy

In this section, I begin by describing the methods I use to answer the first three descriptive questions. I then move into the empirical strategy for the fourth question, which aims to begin to isolate the effect of turnaround principals on school outcomes.

Descriptive analysis. To answer RQ1, which examines the frequency and sustainability of turnaround, I run descriptive statistics showing the frequency of each of the six definitions of turnaround illustrated in Figure 3-2. In RQ2, about the rate of principal turnover in low-performing schools relative to non-low-performing schools, I compare the turnover rates of effective principals in low-performing schools to turnover rates in categories of non-low-performing schools by both definitions of low performing (bottom 10% and bottom 25%) using means, standard deviations, and descriptive regressions (linear probability models). I define the groups according to school performance in the year prior to the principal's first year in a school (i.e., a school that was low-performing the year before a principal began in that school counts as low-performing for all years the principal is there). For this question, turnover is defined as leaving a school, regardless of pathway out. RQ3 extends this analysis by separately examining the pathways out. Specifically, I run descriptives of each pathway out for effective principals of low-performing and non-low-performing schools.

The turnaround principal effect. The ideal approach to estimating the effects of turnaround principals would be to randomly assign effective principals without regard to the type of school in which they had achieved effectiveness, i.e., turnaround principals and other effective principals, to low-performing schools. In reality, a bevy of unobserved factors contribute to principal mobility and school assignments. Some of these factors are voluntary—principals may choose to move from one school to another due to personal or professional reasons, or interest in working in the receiving school's particular context or location. Alternatively, moves may be involuntary or driven by factors endogenous to principal effectiveness—for example, districts or local school boards may shift principals between schools for strategic reasons, or a principal may be compelled to leave a school due to underperformance that may not be observable in administrative data. Given my comparison of interest—in the ideal random assignment study described above, principals assigned to low-performing schools after demonstrating effectiveness in non-low-performing schools—this second set of reasons for leaving is likely to introduce bias into the estimated turnaround principal effect. In particular, principals who shift from a higher performing to a lower performing school may be compelled to move due to underperformance not captured in school achievement data such as community dissatisfaction or a negative school climate. Additionally, school-level factors at the receiving school that led to the principal transition may also confound the estimated effect of turnaround principals if those factors are also associated with the choice of replacement principal.

To that end, there are two sources of endogeneity that I attempt to mitigate in estimating the effect of turnaround principals—unobserved reasons why the receiving school has an open principalship and why the principal is leaving her prior school. In this section, I detail the initial estimation strategy and then describe how I supplement the estimation strategy to address each potential source of endogeneity. To examine the influence of turnaround principals on student achievement growth, I estimate a model that takes the form

$$TestScore_{ijkst} = \beta_0 + \beta_1 TurnaroundPrc_{kt} + \beta_2 EffectivePrc_{kt} + \beta_3 LPS_{st} + \beta_4 TurnaroundPrc_{kt} \times LPS_{st} + \beta_5 EffectivePrc_{kt} \times LPS_{st} + \beta_6 TestScore_{ijt-1} + \beta_7 MissPriorPrcData_{kst} + \theta + \mu + \tau + \alpha \boldsymbol{C}'_{st} + \rho \boldsymbol{P}'_{k} + \phi \boldsymbol{E}'_{kt} + \gamma \boldsymbol{S}'_{it} + \varepsilon_{ijskt} \ , \tag{1}$$

estimating the test score for student *i* in subject *j* under principal *k* at school *s* in year *t*. *TurnaroundPrc* is a binary indicator that takes a value of 1 if the school has a turnaround principal according to the definition examined in the model, *EffectivePrc* is a binary indicator that takes a value of 1 if the school has an effective principal under the same definition but coming from a non-low-performing school, *LPS* is a binary indicator denoting whether the current school was designated as low-performing in the year prior to the principal's first year in the school, $TestScore_{t-1}$ is student *i*'s test score on in exam subject *j* in the prior year, and *MissPriorPrcData* is an indicator that takes a value of 1 if I do not have a value-added score for the principal prior to her first year in school s. θ is a school fixed effect to account for unobserved stable school characteristics that may contribute to a student's performance. μ is a subject fixed effect, with math as the reference category. τ is a year fixed effect to account for unobserved year-to-year changes contributing to student performance. ***C'*** is a vector of time-varying school characteristics that includes average daily membership (linear and squared), per pupil expenditures (linear and squared), school minority percentage, and economically disadvantaged percentage. ***P'*** is a vector of principal characteristics that includes an indicator with a value of 1 for female principals and a set of indicators for principal race with white as the reference category. ***E'*** is a vector of principal experience indicators to allow for a nonlinear effect of principal experience on school outcomes. ***S'*** is a vector of student characteristics that includes an indicator that takes the value of 1 for female students, a set of indicators for student race with white as the reference category, an indicator for whether a student is classified as economically disadvantaged, an indicator for whether a student is an English learner, and an indicator that takes the value of 1 for students who transferred in nonstructurally. ε is an idiosyncratic error term clustered at the school level.

I estimate the same models for the other outcomes, omitting lagged test score and subject for all other outcomes, and replacing the vector of student covariates with teacher covariates in the model predicting teacher turnover. Because teacher turnover and chronic student absenteeism are binary outcomes, the estimates for these two effects are from linear probability models in which the estimates of interest represent the change in probability associated with having a turnaround principal, conditional on all other variables in the model. The teacher turnover model replaces the student covariates with a vector of teacher covariates that includes an indicator that takes the value of 1 for female teachers, a set of indicators for teacher race with white as the reference category, a set of indicators for teacher education level with bachelor's as the reference category, and teaching experience (linear and squared).

The linear combination of β_0, β_1, β_3, and β_4 represents the estimate of having a turnaround principal in a low-performing school, while the linear combination of β_0, β_2, β_3, and β_5 represents the effect of having an effective principal from a non-low-performing school in a low-performing school. The difference between those two linear combinations provides the answer to RQ4, which asks whether turnaround principals are more effective than effective principals coming from non-low-performing schools. If an F-test shows the difference between the linear combination of β_1 and β_4 and the linear combination of β_2 and β_5 is significantly different from zero, I can conclude that there is a differential effect of turnaround principals and other effective principals. If there is no significant difference but the two linear combinations are both positive and significant, then low-performing schools benefit from having principals with demonstrated effectiveness in prior schools—regardless of whether those schools were low performing. This latter finding would suggest the skillset that makes a principal effective is transferable across school types, and the notion of a turnaround principal may not be substantively different from that of an effective principal. Finally, on their own, the estimates on β_1 and β_2, provide information about the persistent effect of turnaround and other effective principals, respectively, relative to principals who have not demonstrated prior effectiveness. Positive coefficients would provide evidence for the portability of principal effectiveness.

These estimates still may be confounded by the school- and principal-level sources of endogeneity described above. At the school level, a particular concern is a pre-transition dip that led to the principal transition. For example, a district may dismiss a principal due to poor performance or an outgoing principal may withdraw effort prior to her departure, leading to a dip in school outcomes. These scenarios will introduce bias if the pre-transition dip is more or less likely to occur in schools that ultimately recruit new principals coming from low-performing schools than schools that recruit principals from non-low-performing schools; for example, schools that hire principals from higher performing schools may have resources such as more experienced teachers or stronger district leadership that increase their ability to recruit effective principals. To that end, the pre-transition dip may be more frequent in schools that recruit from low-performing schools if, for example, the applicant pool is sparser for schools in decline than more stable schools. To control for the possibility of a pre-transition dip, I add a leading indicator that takes the value of 1 in the year prior to the principal transition, with the new model taking the form

$$TestScore_{ijkst} = \beta_0 + \beta_1 TurnaroundPrc_{kt} + \beta_2 EffectivePrc_{kt} + \beta_3 LPS_{st} + \beta_4 TurnaroundPrc_{kt} \times LPS_{st} + \beta_5 EffectivePrc_{kt} \times LPS_{st} + \beta_6 TestScore_{ijt-1} + \beta_7 MissPriorPrcData_{kst} + \beta_8 PreTransition_{st+1} + \theta + \mu + \tau + \alpha \boldsymbol{C'}_{st} + \rho \boldsymbol{P}'_k + \phi \boldsymbol{E}'_{kt} + \gamma \boldsymbol{S'}_{it} + \varepsilon_{ijkst} \ , \quad (2)$$

The coefficient on this indicator, β_8, will absorb a one-year performance dip that occurs prior to the transition, though it will treat all pre-transition dips equally. If the size of the pre-transition dip is associated with both student achievement and the choice of replacement principal, this indicator will not mitigate bias resulting from regression to the mean in the year following the transition.

Of particular importance to all of these estimates is the requirement that principals move from one school to another in order to estimate the turnaround principal effect. I classify a school as having a turnaround principal if its principal was effective in a prior low-performing school. A school whose principal meets the turnaround principal definition during the tenure that includes year *t* would not be classified as having a turnaround principal. To that end, the estimated turnaround principal and other effective principal effects are based on principals in their second school or beyond during the study period.

Finally, while I cannot fully account for the possibility that the principals I identify as effective do not have unobservable traits making them less effective than principals moving from one low-performing school to another, I test the equivalence of these two groups on lagged relevant measures that may signal principal effectiveness but are not included in my definitions of turnaround principals. These include principal NCEES ratings on each of the seven dimensions as well as the median principal rating, school-level teacher turnover rates, and four TWC dimensions that may signal principal effectiveness: school leadership, teacher leadership, community support and involvement, and managing student conduct. To measure differences in a principal's influence on a school rather than capturing the differences in these measures that are correlated with school performance, I use deviations from the school-level mean teacher turnover rate and TWC dimensions rather than one-year levels. I conduct these balance checks using data from the principal's last year in the school in which they most recently met the criteria for a turnaround or effective principal. In cases in which the principal's last year is a non-TWC year, I draw TWC measures from the year prior. I exclude principals from these balance checks whose last year is a non-TWC year but who only spent a single year in the school.

Results

RQ1: How frequently are principals successful in low-performing schools and how often do they sustain that success over multiple years?

As shown in Table 3-2 above, about one-third percent of principals in low-performing schools in any given year are effective, as measured by principal value added. About half are effective in at least one of their first two years in that school, as shown in Table 3-3. Fewer are able to sustain that effectiveness over multiple years, with Table 3-4 showing that of principals who remained in their low-performing schools for at least three years, just over 30 percent were effective for at least two of those years and 22 to 24 percent were effective for three.

RQ2: Do successful principals of low-performing schools leave at higher rates than successful principals of non-low-performing schools?

Descriptively, successful principals leave low-performing schools at higher rates than they leave non-low-performing schools. For example, Table 3-7 shows about 18 percent of principals meeting the most liberal definition of effectiveness leave non-low-performing schools (defined as the bottom 10%) while about 21 percent leave low-performing schools. Under the most conservative definition of effectiveness, 12 percent leave non-low-performing schools while 17 percent leave low-performing schools.

Table 3-7.1 Attrition rates by effectiveness definition and school low-performing status (LP=bottom 10%)

Panel A. All principals

	Ineffective		Def A		Def B		Def C		Def D		Def E		Def F	
	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N
Non-LPS	0.178	14757	0.188	5253	0.165	3938	0.142	4045	0.144	3948	0.121	3019	0.122	2992
	(0.382)		(0.391)		(0.371)		(0.349)		(0.352)		(0.326)		(0.327)	
LPS	0.210	1287	0.242	438	0.230	404	0.202	381	0.204	378	0.171	258	0.171	258
	(0.407)		(0.429)		(0.421)		(0.402)		(0.403)		(0.377)		(0.377)	
N	16044		5691		4342		4426		4326		3277		3250	

Panel B. Principals with 3+ years in school

	Ineffective		Def A		Def B		Def C		Def D		Def E		Def F	
	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N
Non-LPS	0.120	11499	0.137	4158	0.122	3125	0.121	3518	0.123	3421	0.121	3019	0.122	2992
	(0.325)		(0.344)		(0.328)		(0.326)		(0.329)		(0.326)		(0.327)	
LPS	0.120	900	0.174	305	0.168	286	0.167	305	0.169	302	0.171	258	0.171	258
	(0.325)		(0.380)		(0.374)		(0.374)		(0.375)		(0.377)		(0.377)	
N	12399		4463		3411		3823		3723		3277		3250	

School-by-year observations. Standard deviations in parentheses.

Table 3-7.2 Attrition rates by effectiveness definition and school low-performing status (LP=bottom 25%)

Panel A. All principals

	Ineffective		Def A		Def B		Def C		Def D		Def E		Def F	
	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N
Non-LPS	0.176	12995	0.186	4587	0.161	3387	0.139	3485	0.141	3410	0.119	2594	0.119	2583
	(0.381)		(0.389)		(0.367)		(0.346)		(0.348)		(0.324)		(0.324)	
LPS	0.199	3049	0.220	1104	0.208	955	0.176	941	0.181	916	0.148	683	0.151	667
	(0.399)		(0.415)		(0.406)		(0.381)		(0.385)		(0.355)		(0.359)	
N	16044		5691		4342		4426		4326		3277		3250	

Panel B. Principals with 3+ years in school

	Ineffective		Def A		Def B		Def C		Def D		Def E		Def F	
	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N	Prop	N
Non-LPS	0.120	10188	0.135	3652	0.118	2700	0.119	3036	0.120	2961	0.119	2594	0.119	2583
	(0.325)		(0.342)		(0.323)		(0.323)		(0.325)		(0.324)		(0.324)	
LPS	0.123	2211	0.158	811	0.156	711	0.149	787	0.154	762	0.148	683	0.151	667
	(0.329)		(0.365)		(0.363)		(0.356)		(0.361)		(0.355)		(0.359)	
N	12399		4463		3411		3823		3723		3277		3250	

School-by-year observations. Standard deviations in parentheses.

However, principals leave low-performing schools at higher rates than non-low-performing schools regardless of principal effectiveness. Table 3-8 highlights three relevant takeaways. First, principals are more likely to leave low-performing schools than non-low-performing schools. Second, effective principals—by all but my most liberal definition—are less likely to leave their schools, on average, than ineffective principals. However, third, as shown by the insignificant coefficients on the interaction term, turnaround principals are not statistically significantly more likely to turn over than other effective principals.

Table 3-8. Probability of attrition by school low-performing status and principal effectiveness by definition

	Def A	Def B	Def C	Def D	Def E	Def F
Effective	0.010	-0.019**	-0.048***	-0.045***	-0.070***	-0.069***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
LPS	0.032**	0.030**	0.032**	0.033**	0.035***	0.035***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Effective x LPS	0.022	0.035	0.028	0.027	0.014	0.013
	(0.022)	(0.023)	(0.023)	(0.024)	(0.027)	(0.027)
Constant	0.178***	0.184***	0.190***	0.189***	0.191***	0.191***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	21735	21735	21735	21735	21735	21735

Coefficients from descriptive linear probability models. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

RQ3: Where do principals with demonstrated prior success in low-performing schools ("turnaround principals") go after leaving the formerly low performing school?

The modal pathway out of a school is to another principalship within the same district, with 39.5% of all principals who turned over taking this pathway. Table 3-9 shows that turnaround principals made these intradistrict moves in 39 to 44 percent of turnovers, depending on the definition. Effective principals at non-low-performing schools made these intradistrict moves at even higher rates than effective principals low-performing schools under all but the most liberal definition of effectiveness.

Table 3-9. Pathway out by principal effectiveness and low-performing status

Panel A. Low performing as bottom 10%

	Ineffective	Def A	Def B	Def C	Def D	Def E	Def F
Intradistrict mover							
Low performing	30.7	41.5	39.8	39.0	39.0	43.2	43.2
Not low performing	39.7	41.1	42.5	43.2	43.2	43.0	43.1
Interdistrict mover							
Low performing	10.4	7.5	6.5	6.5	6.5	4.5	4.5
Not low performing	5.0	7.7	9.1	7.7	7.7	6.0	6.0
Promotion							
Low performing	13.7	21.7	23.7	20.8	20.8	22.7	22.7
Not low performing	17.3	18.8	19.5	20.4	20.4	21.6	21.4
Demotion							
Low performing	28.1	11.3	12.9	15.6	15.6	11.4	11.4
Not low performing	17.6	15.8	15.5	13.8	13.7	13.7	13.7
Leaver							
Low performing	17.0	17.9	17.2	18.2	18.2	18.2	18.2
Not low performing	20.5	16.6	13.4	15.0	15.1	15.6	15.7

Panel B. Low performing as bottom 25%

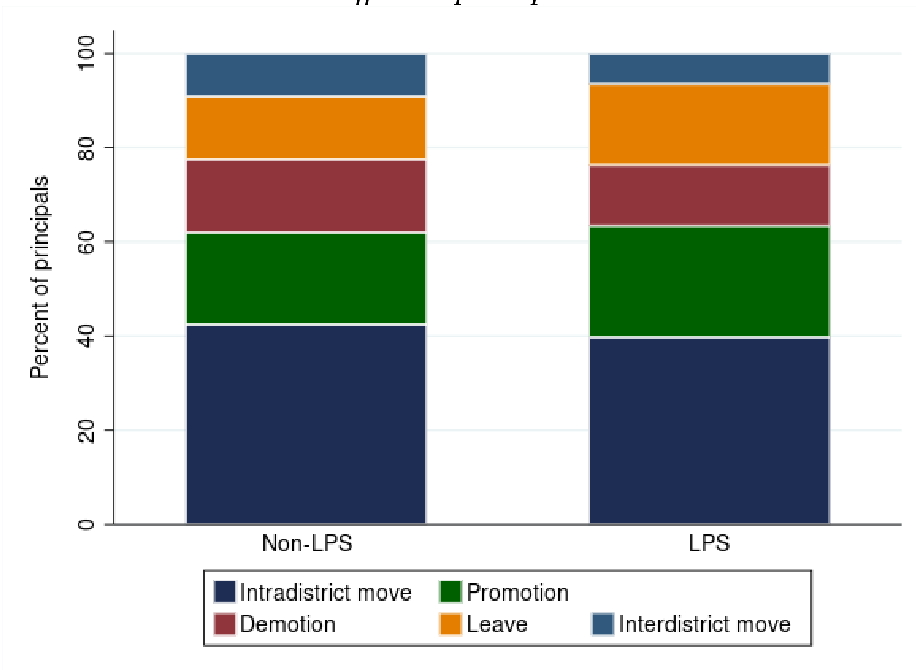
	Ineffective	Def A	Def B	Def C	Def D	Def E	Def F
Intradistrict mover							
Low performing	30.7	44.4	41.7	40.4	40.4	42.6	42.6
Not low performing	39.7	40.2	42.3	43.5	43.5	43.2	43.3
Interdistrict mover							
Low performing	10.4	7.0	7.0	6.6	6.6	5.9	5.9
Not low performing	5.0	7.9	9.4	7.8	7.9	5.8	5.9
Promotion							
Low performing	13.7	16.9	19.6	19.9	19.9	19.8	19.8
Not low performing	17.3	19.7	20.2	20.6	20.6	22.4	22.1
Demotion							
Low performing	28.1	14.8	15.6	15.1	15.1	15.8	15.8
Not low performing	17.6	15.5	15.1	13.6	13.5	12.7	12.7
Leaver							
Low performing	17.0	16.9	16.1	18.1	18.1	15.8	15.8
Not low performing	20.5	16.7	13.1	14.4	14.6	15.9	16.0

While intradistrict moves were the modal pathway out for turnaround principals, turnaround principals left and were promoted at higher rates than other effective principals. By contrast, principals who were ineffective in low-performing schools were demoted more often than principals who were ineffective in non-low-performing schools. Figure 3-3 shows these pathways out graphically, first for turnaround and effective principals defined using Definition B[24] (Panel A), and then for ineffective principals in low-performing and non-low-performing schools (Panel B).

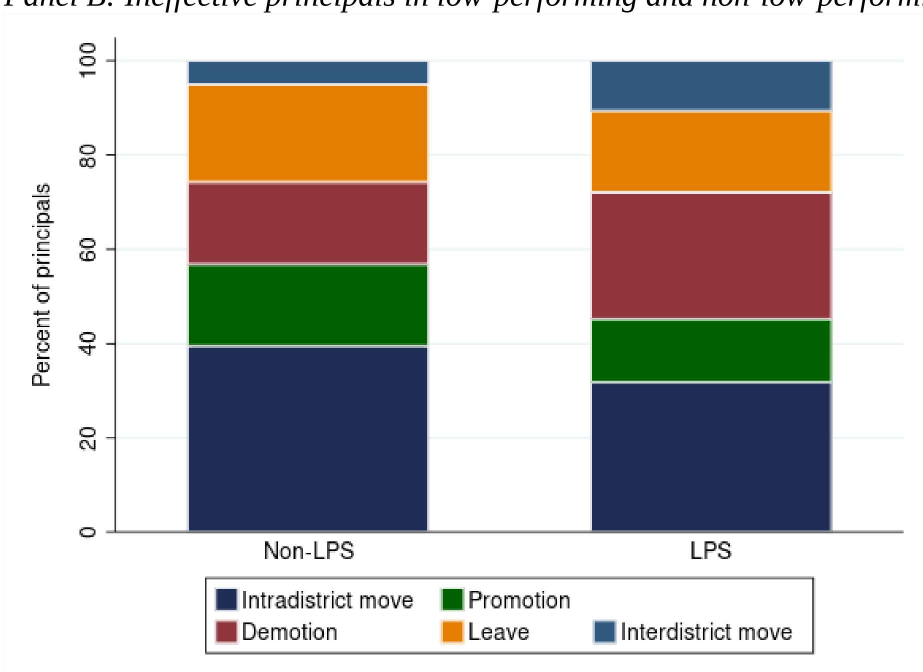
[24] Pathways out were similar for each of the six definitions.

Figure 3-3. Pathway out by principal effectiveness definition B and school low-performing status (LP=bottom 10%)

Panel A: Turnaround and effective principals



Panel B: Ineffective principals in low-performing and non-low-performing schools



Of the turnaround principals who move to another principalship, most move to other low-performing schools, as shown in Table 3-10. These mover pathways look similar to pathways of ineffective principals, suggesting that principals of lower performing schools tend to move to other low-performing schools, regardless of prior success.

Table 3-10. Pathway out for movers by turnaround principal category

Panel A. Low performing as bottom 10%

	0 to <5th percentile		5 to <10th percentile		10 to < 25th percentile	
	Percent	Sum	Percent	Sum	Percent	Sum
Ineffective	53.4	87	44.8	73	1.8	3
Def A	70.0	7	30.0	3	0.0	0
Def B	66.7	6	33.3	3	0.0	0
Def C	57.1	4	42.9	3	0.0	0
Def D	57.1	4	42.9	3	0.0	0
Def E	80.0	4	20.0	1	0.0	0
Def F	80.0	4	20.0	1	0.0	0

Panel B. Low performing as bottom 25%

	0 to <5th percentile		5 to <10th percentile		10 to < 25th percentile	
	Percent	Sum	Percent	Sum	Percent	Sum
Ineffective	23.2	76	19.3	63	57.5	188
Def A	30.2	19	23.8	15	44.4	28
Def B	27.7	13	27.7	13	0.0	0
Def C	25.8	8	29.0	9	0.0	0
Def D	25.8	8	29.0	9	0.0	0
Def E	26.3	5	26.3	5	0.0	0
Def F	26.3	5	26.3	5	0.0	0

Sample restricted to mover principals in the first year of their new school. Proficiency rate of receiving school based on prior year proficiency. An additional 22 principals meeting at least the minimum definition of turnaround principal moved to either a school that opened that year or with no tested grades and therefore no proficiency rate.

RQ4: Are these turnaround principals more effective at leading low-performing schools than principals who have successfully led non-low-performing schools?

Student achievement. I do not find evidence that students in low-performing schools with either turnaround or other effective principals have higher test score gains than students without principals who do not meet these definitions. Table 3-11 presents the results. The coefficients on the turnaround principal variable are consistently small and nonsignificant, which indicates that students do not make larger achievement gains in years when their school is helmed by a turnaround principal than in years without a turnaround principal. The nonsignificant interactions between turnaround principal and low-performing school indicate that students do not fare better under turnaround principals in low-performing schools than in non-low-performing schools. The linear combination of these two coefficients is .008 to .009 when defining low performing as the bottom 10 percent of schools and .02 to .021 when defining low performing as the bottom 25 percent, but neither is statistically significant, which suggests that there is no effect of turnaround principals on student achievement in low-performing schools. The corresponding coefficient estimates on effective principals and the interaction between effective principal and low-performing school similarly find no evidence that students with effective principals, by these definitions, fare better than their counterparts—either overall or in low-performing schools.

I also do not find evidence that students in low-performing schools led by turnaround principals make larger achievement gains than their counterparts in low-performing schools led by other effective principals. In particular, the non-significant F-statistics at the bottom of the table indicate that the linear combination of turnaround principal and turnaround principal X LPS is not significantly different from the linear combination of effective principal and effective principal X LPS. These null findings are consistent across both math and reading (shown separately in appendix Table A-3-1 and Table A-3-2, respectively).

Table 3-11. Turnaround and effective principal effects on student achievement

	LP as bottom 10%		LP as bottom 25%	
	(1)	(2)	(3)	(4)
Turnaround principal	0.016	0.015	0.002	0.001
	(0.016)	(0.016)	(0.011)	(0.011)
Effective principal	0.006	0.005	0.005	0.005
	(0.005)	(0.005)	(0.006)	(0.006)
LPS	-0.004	-0.004	-0.008	-0.008
	(0.007)	(0.007)	(0.005)	(0.005)
Turnaround principal X LPS	-0.007	-0.007	0.019	0.019
	(0.028)	(0.028)	(0.016)	(0.016)
Effective principal X LPS	0.003	0.003	-0.004	-0.004
	(0.015)	(0.016)	(0.011)	(0.011)
TP/EP missing	-0.007*	-0.006	-0.007*	-0.007*
	(0.003)	(0.003)	(0.003)	(0.003)
Pre-transition year		-0.003*		-0.004*
		(0.002)		(0.002)
Constant	0.166***	0.164***	0.165***	0.163***
	(0.031)	(0.031)	(0.031)	(0.031)
F (TP=EP)	0.00	0.00	1.47	1.34
p-value (EP=TP)	0.999	0.987	0.226	0.247
R^2	0.704	0.704	0.704	0.704
Obs	9081029	9081029	9081029	9081029

All models include school, principal, and student covariates, and school, year, and subject fixed effects. Standard errors clustered at the school level. F-statistics test the equivalence of the linear combination of "Turnaround principal" + "Turnaround principal x LPS" with the linear combination of "Effective principal" + "Effective principal x LPS." * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Models 2 and 4 show that while principal turnover does not seem to suppress the turnaround or effective principal effect, it is associated with a .003 to .004 standard deviation dip in student achievement growth.

Teacher mobility. I do not find evidence that turnaround principals or other effective principals are associated with decreased teacher turnover in low-performing schools, as Table

3-12 shows. In fact, the main effect in Table 3-12 shows that teacher turnover is higher under turnaround principals, on average, than non-turnaround principals—the models controlling for principal turnover find that the probability of teacher turnover is 2.5 to 2.6 percentage points higher when a school has a turnaround principal than when a school has a principal who does not meet the effectiveness definition. However, in the models defining low performing as the bottom 25 percent of schools, the significant interaction term between turnaround principal and the low-performing school variable indicates that turnaround principals produce significantly lower teacher turnover in low-performing schools than they do in non-low-performing schools. The non-significant F-statistics indicate that turnaround principals and effective principals do not have a significantly different effect on teacher turnover in low-performing schools.

Table 3-12. Turnaround and effective principal effects on teacher turnover

	LP as bottom 10%		LP as bottom 25%	
	(1)	(2)	(3)	(4)
Turnaround principal	0.018	0.026*	0.018	0.025*
	(0.010)	(0.010)	(0.011)	(0.011)
Effective principal	-0.004	0.001	-0.002	0.002
	(0.005)	(0.005)	(0.005)	(0.005)
LPS	0.003	0.006	0.003	0.006
	(0.007)	(0.006)	(0.005)	(0.005)
Turnaround principal X LPS	-0.017	-0.018	-0.038*	-0.038*
	(0.033)	(0.033)	(0.018)	(0.018)
Effective principal X LPS	-0.021	-0.021	-0.012	-0.013
	(0.014)	(0.014)	(0.013)	(0.013)
TP/EP missing	-0.011**	-0.014***	-0.010**	-0.014***
	(0.003)	(0.003)	(0.003)	(0.003)
Pre-transition year		0.023***		0.023***
		(0.002)		(0.002)
Constant	0.049	0.063	0.047	0.061
	(0.084)	(0.084)	(0.084)	(0.084)
F (TP=EP)	0.47	0.54	0.08	0.01
p-value (EP=TP)	0.493	0.462	0.779	0.904
R^2	0.051	0.052	0.051	0.052
Obs	580995	580995	580995	580995

All models include school, principal, and student covariates, and school, year, and subject fixed effects. Standard errors clustered at the school level. F-statistics test the equivalence of the linear combination of “Turnaround principal” + “Turnaround principal x LPS” with the linear combination of “Effective principal” + “Effective principal x LPS.” * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Again, principal turnover does not suppress the turnaround or effective principal effects, but it is associated with a higher probability of teacher turnover. In particular, these results show that the probability of turnover is 2.3 percentage points higher in the year of a principal turnover.

Chronic absenteeism. Neither turnaround principals or other effective principals appear to reduce chronic absenteeism more than other principals, as Table 3-13 shows. The coefficient estimates on the turnaround principal indicator are consistently negative but nonsignificant. It is possible from these estimates that turnaround principals reduce chronic absenteeism but that I am unable to detect such a small effect size. The linear combination of turnaround principal and the interaction term is nonsignificant and very close to zero, providing no evidence that turnaround principals reduce chronic absenteeism in low-performing schools. The nonsignificant F-statistics show there is no significant difference between the effect of turnaround principals and effective principals in low-performing schools on chronic absenteeism.

Table 3-13. Turnaround and effective principal effects on chronic absenteeism

	LP as bottom 10%		LP as bottom 25%	
	(1)	(2)	(3)	(4)
Turnaround principal	-0.006	-0.005	-0.006	-0.006
	(0.004)	(0.004)	(0.004)	(0.004)
Effective principal	-0.001	-0.001	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
LPS	0.011***	0.011***	0.005**	0.005**
	(0.003)	(0.003)	(0.002)	(0.002)
Turnaround principal X LPS	0.002	0.002	0.003	0.003
	(0.012)	(0.012)	(0.008)	(0.008)
Effective principal X LPS	0.003	0.003	0.001	0.001
	(0.006)	(0.006)	(0.004)	(0.004)
TP missing	0.001	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Principal turnover		0.002***		0.002***
		(0.001)		(0.001)
Constant	-0.019	-0.018	-0.018	-0.017
	(0.012)	(0.012)	(0.012)	(0.012)
F (TP=EP)	0.19	0.18	0.27	0.22
p-value (EP=TP)	0.663	0.675	0.603	0.637
R^2	0.072	0.072	0.072	0.072
Obs	6539747	6539747	6539747	6539747

All models include school, principal, and student covariates, and school, year, and subject fixed effects. Standard errors clustered at the school level. F-statistics test the equivalence of the linear combination of "Turnaround principal" + "Turnaround principal x LPS" with the linear combination of "Effective principal" + "Effective principal x LPS."

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

While I find no effect of turnaround and effective principals on chronic absenteeism, two significant findings are of note. First, chronic absenteeism is slightly higher in low-performing schools (.5 to 1.1 percentage points). Second, chronic absenteeism increases by about .2 percentage points in the year of a principal turnover.

Robustness checks

Turnaround principals and other effective principals appear to differ in their effectiveness at prior schools as measured by teachers perceptions of school leadership and community support and involvement. In particular, turnaround principals who move to new schools appear to have been less effective on these measures at their prior schools than other effective principals when defining low performing as the bottom quartile of schools (Table A-3-3). It is possible that some of these null effects may be driven in part by baseline differences in effectiveness between the two groups. However, it is also possible that these measures, in part, capture dissatisfaction from teachers who are not supportive of the changes a turnaround principal creates in a school. In this latter scenario, the differences would not necessarily bias the estimates of turnaround principals on student achievement. It could, however, explain the higher teacher turnover under these principals.

Discussion

By and large, I do not find evidence that prior success in a low-performing school is predictive of improved student or teacher outcomes. There are four broad explanations for these results. First, it may be that principal effectiveness does not translate from one school to the next. There is some evidence that principal improvement in one school does not carry over to the next school (Bartanen, 2019); my findings may provide evidence that principal effectiveness does not carry over either. Second, there may be heterogeneity that my definitions are unable to parse; for example, the mechanism through which principals raise test scores in their original schools might matter for the portability of their effectiveness. Third, principal value-added might not be the right tool for identifying turnaround and effectiveness principals—or principal value-added on test scores may represent only one dimension of a principal's effectiveness. Existing research suggests principal effects are likely to filter through school-level factors even more proximal to the principalship, such as relational trust within the school, teacher commitment and job satisfaction, and school culture (Hanselman et al., 2016; Heck et al., 1990; Hulpia et al., 2009; Mascall & Leithwood, 2010). The portability of principal effectiveness from one low-performing school to another may depend on factors more proximal to the principal's influence, such as teacher turnover or school climate. Finally, the effect of a turnaround principal may take time to materialize. With additional years of data or a definition that classified a higher number of principals as turnaround principals, it may be possible to estimate separate effects for the first, second, and third year of a new principal. Given the small number principals my definition identified as turnaround principals, I did not have large enough cell sizes to test this hypothesis.

Additionally, all of these estimates contain quite a bit of noise for two reasons. First, because the principal value-added measures rely on connected networks that in many cases contain very few principals, they may misclassify principals, adding to the measurement error of the turnaround and effective principal indicators. Second, school fixed effects require that schools have at least one principal who entered while the school was low-performing and one who entered while the school was not low-performing. The estimates on the low-performing school indicator and the interactions with that indicator therefore are based on a comparison that may be driven by other confounding school-level factors. Finally, although I include an indicator for principals for whom I cannot observe sufficient prior performance to code as a turnaround principal, the turnaround and effective principal indicators still do not include principals who may be effective even though I do not observe their prior effectiveness because they came from an out-of-state or private school, a school in a disconnected network, a school without tested grades, met the effectiveness criteria in a school prior to the study period. As a result, the turnaround and effective principal indicators likely capture only a subset of principals who may fall into this categories. A more complete picture of each principal's prior experience would provide more power to detect an effect of turnaround principals.

In sum, the findings from this study do not provide evidence that high value-added in a low-performing school predicts future effectiveness. However, it remains unclear whether there truly are turnaround principals with a particular set of skills conducive to turning around low-performing schools. Future research could investigate this question by drawing from longer panels of data to track principals across schools, or use varying definitions of principal effectiveness that draw from multiple dimensions of principal effectiveness beyond student test scores. By honing in on a more nuanced operationalization of turnaround principals, future research could move toward answering the question of whether there are turnaround principals.

References

Aladjem, D. K., Birman, B. F., Orland, M., Harr-Robins, J., Heredia, A., Parrish, T. B., & Ruffini, S. J. (2010). *Achieving Dramatic School Improvement: An Exploratory Study*. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. https://eric.ed.gov/?id=ED526783

Bartanen, B. (2019). *Identifying Principal Improvement*. https://doi.org/10.26300/KCRZ-TM19

Bartanen, B. (2020). Principal Quality and Student Attendance. *Educational Researcher*, *49*(2), 101–113. https://doi.org/10.3102/0013189X19898702

Battle, D. (2010). Principal Attrition and Mobility: Results from the 2008-09 Principal Follow-Up Survey. First Look. NCES 2010-337. *National Center for Education Statistics*.

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, *41*(4), 904–919. https://doi.org/10.1016/j.ssresearch.2012.03.003

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review*, *95*(2), 166–171.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals. *National Bureau of Economic Research Working Paper Series*, *No. 17803*. https://doi.org/10.3386/w17803

Branch, G. F., Hanushek, E., & Rivkin, S. (2009). Estimating Principal Effectiveness. Working Paper 32. *National Center for Analysis of Longitudinal Data in Education Research*.

Burkhauser, S. (2017). How Much Do School Principals Matter When It Comes to Teacher Working Conditions? *Educational Evaluation and Policy Analysis*, *39*(1), 126–145. https://doi.org/10.3102/0162373716668028

Calkins, A., Guenther, W., Belfiore, G., & Lash, D. (2007). *The Turnaround Challenge: Why America's Best Opportunity to Dramatically Improve Student Achievement Lies in Our Worst-Performing Schools*. Mass Insight Education. https://eric.ed.gov/?id=ED538298

Carlson, D., & Lavertu, S. (2018). School Improvement Grants in Ohio: Effects on Student Achievement and School Administration. *Educational Evaluation and Policy Analysis*, 0162373718760218. https://doi.org/10.3102/0162373718760218

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *The American Economic Review*, *104*(9), 2633–2679.

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is School Value Added Indicative of Principal Quality? *Education Finance and Policy*, *11*(3), 283–309. https://doi.org/10.1162/EDFP_a_00184

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24*(4), 377–392. https://doi.org/10.1016/j.econedurev.2004.06.008

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*(1), 92–109. https://doi.org/10.1016/j.econedurev.2011.09.001

Copeland, G., & Neeley, A. (2013). *Identifying Competencies and Actions of Effective Turnaround Principals: Southeast Comprehensive Center Briefing Paper*. The Southeast Comprehensive Center. http://www.sedl.org/pubs/catalog/items/secc05.html

DeAngelis, K. J., & White, B. R. (2011). *Principal Turnover in Illinois Public Schools, 2001-2008. Policy Research: IERC 2011-1*. Illinois Education Research Council. https://eric.ed.gov/?id=ED518191

Dhuey, E., & Smith, J. (2018). How school principals influence student learning. *Empirical Economics*, *54*(2), 851–882. https://doi.org/10.1007/s00181-017-1259-9

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in Texas. *The Georgetown Public Policy Review*, 21–39.

Dodman, S. (2014). A Vivid Illustration of Leadership: Principal's Actions Propel Struggling School's Turnaround. *Journal of Staff Development*, *35*(1), 56–58.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., & Giffin, J. (2017). *School Improvement Grants: Implementation and Effectiveness. NCEE 2017-4013*. National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED572215

Duke, D. L. (2004). The Turnaround Principal: High-Stakes Leadership. *Principal*, *84*(1), 12–23.

Duke, D. L., & Landahl, M. (2011). "Raising Tests Scores Was the Easy Part": A Case Study of the Third Year of School Turnaround. *International Studies in Educational Administration (Commonwealth Council for Educational Administration & Management (CCEAM))*, *39*(3).

Duke, D. L., & Salmonowicz, M. (2010). Key Decisions of a First-year 'Turnaround' Principal. *Educational Management Administration & Leadership*, *38*(1), 33–58. https://doi.org/10.1177/1741143209345450

Elmore, R. F. (2000). *Building a New Structure for School Leadership*. Albert Shanker Institute. https://eric.ed.gov/?id=ED546618

Finnigan, K. S., & Stewart, T. J. (2009). Leading Change under Pressure: An Examination of Principal Leadership in Low-Performing Schools. *Journal of School Leadership*, *19*(5), 586–621. https://doi.org/10.1177/105268460901900504

Fuller, E. J., & Young, M. D. (2009). *Tenure and retention of newly hired principals in Texas*.

Gates, S. M., Ringel, J. S., Santibañez, L., Guarino, C., Ghosh-Dastidar, B., & Brown, A. (2006). Mobility and turnover among school principals. *Economics of Education Review*, *25*(3), 289–302. http://dx.doi.org/10.1016/j.econedurev.2005.01.008

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students. *Educational Researcher*, *44*(5), 293–307. https://doi.org/10.3102/0013189X15592622

Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, *113*(11), 2552–2585.

Grissom, J. A. (2012). Revisiting the Impact of Participative Decision Making on Public Employee Retention: The Moderating Influence of Effective Managers. *The American Review of Public Administration*, *42*(4), 400–418. https://doi.org/10.1177/0275074011404209

Grissom, J. A., & Bartanen, B. (2019). Strategic Retention: Principal Effectiveness and Teacher Turnover in Multiple-Measure Teacher Evaluation Systems. *American Educational Research Journal*, *56*(2), 514–555. https://doi.org/10.3102/0002831218797931

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3–28.

Grissom, J. A., & Loeb, S. (2011). Triangulating Principal Effectiveness: How Perspectives of Parents, Teachers, and Assistant Principals Identify the Central Importance of Managerial Skills. *American Educational Research Journal*, *48*(5), 1091–1123. https://doi.org/10.3102/0002831211402663

Grissom, J. A., Loeb, S., & Mitani, H. (2015). Principal time management skills: Explaining patterns in principals' time use, job stress, and perceived effectiveness. *Journal of Educational Administration*, *53*(6), 773–793. https://doi.org/10.1108/JEA-09-2014-0117

Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, *76*(2), 173–208.

Hallinger, P., & Heck, R. H. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*(2), 157–191. https://doi.org/10.1080/0924345980090203

Hanselman, P., Grigg, J., Bruch, S. K., & Gamoran, A. (2016). The Consequences of Principal and Teacher Turnover for School Social Resources. In *Family Environments, School Resources, and Educational Outcomes* (Vol. 19, pp. 49–89). Emerald Group Publishing Limited. https://doi.org/10.1108/S1479-353920150000019004

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why Public Schools Lose Teachers. *The Journal of Human Resources*, *39*(2), 326–354. https://doi.org/10.2307/3559017

Harris, A. (2002). Effective Leadership in Schools Facing Challenging Contexts. *School Leadership & Management*, *22*(1), 15–26. https://doi.org/10.1080/13632430220143024a

Heck, R. H., Larsen, T. J., & Marcoulides, G. A. (1990). Instructional Leadership and School Achievement:Validation of a Causal Model. *Educational Administration Quarterly*, *26*(2), 94–125. https://doi.org/10.1177/0013161X90026002002

Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, *62*, 302–320. https://doi.org/10.1016/j.econedurev.2017.08.001

Henry, G. T., Guthrie, J. E., & Townsend, L. W. (2015). *Outcomes and Impacts of North Carolina's Initiative to Turn Around the Lowest-Achieving Schools*. http://cerenc.org/wp-content/uploads/2015/09/ES-FINAL-Final-DST-Report-9-3-15.pdf

Henry, G. T., & Harbatkin, E. (2018, November 8). *The Next Generation of School Reform: Improving the Lowest Performing Schools without Disrupting the Status Quo*. 2018 APPAM Fall Research Conference: Evidence for Action: Encouraging Innovation and Improvement, Washington, D.C.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., & Redding, S. (2008). *Turning Around Chronically Low-Performing Schools: A Practice Guide* (NCEE 2008-4020; IES Practice Guide). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/PracticeGuide.aspx?sid=7

Hochbein, C. (2012). Relegation and Reversion: Longitudinal Analysis of School Turnaround and Decline. *Journal of Education for Students Placed at Risk (JESPAR)*, *17*(1–2), 92–107. https://doi.org/10.1080/10824669.2012.636728

Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's Time Use and School Effectiveness. *American Journal of Education*, *116*(4), 491–523. https://doi.org/10.1086/653625

Huberman, M., Parrish, T., Hannan, S., Arellanes, M., & Shambaugh, L. (2011). *Turnaround Schools in California: Who Are They and What Strategies Do They Use?* California Comprehensive Center at WestEd. https://eric.ed.gov/?id=ED592859

Hulpia, H., Devos, G., & Rosseel, Y. (2009). The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organizational commitment. *School Effectiveness and School Improvement*, *20*(3), 291–317. https://doi.org/10.1080/09243450902909840

Jacob, B. A., Lefgren, L., & Sims, D. P. (2010). The Persistence of Teacher-Induced Learning. *Journal of Human Resources*, *45*(4), 915–943. https://doi.org/10.3368/jhr.45.4.915

Jacobson, S. L., Giles, C., Ylimaki, R., & Johnson, L. (2005). Successful leadership in challenging US schools: Enabling principles, enabling schools. *Journal of Educational Administration*, *43*(6), 607–618. https://doi.org/10.1108/09578230510625700

Kafka, J. (2009). The Principalship in Historical Perspective. *Peabody Journal of Education*, *84*(3), 318–330. https://doi.org/10.1080/01619560902973506

Leithwood, K., & Jantzi, D. (1990). Transformational Leadership: How Principals Can Help Reform School Cultures. *School Effectiveness and School Improvement*, *1*(4), 249–280. https://doi.org/10.1080/0924345900010402

Leithwood, K., Louis, K. S., Wahlstrom, K., Anderson, S., Mascall, B., & Gordon, M. (2009). How Successful Leadership Influences Student Learning: The Second Installment of a Longer Story. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second International Handbook of Educational Change* (pp. 611–629). Springer Netherlands. https://doi.org/10.1007/978-90-481-2660-6_35

Loeb, S., Kalogrides, D., & Horng, E. L. (2010). Principal Preferences and the Uneven Distribution of Principals Across Schools. *Educational Evaluation and Policy Analysis*, *32*(2), 205–229.

Mao, M. X., Whitsett, M. D., & Mellor, L. T. (1997). *Student Mobility, Academic Performance, and School Accountability*. https://eric.ed.gov/?id=ED409380

Mascall, B., & Leithwood, K. (2010). Investing in Leadership: The District's Role in Managing Principal Turnover. *Leadership and Policy in Schools*, *9*(4), 367–383. https://doi.org/10.1080/15700763.2010.493633

McGee, G. W. (2004). Closing the Achievement Gap: Lessons From Illinois' Golden Spike High-Poverty High-Performing Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *9*(2), 97–125. https://doi.org/10.1207/s15327671espr0902_2

Meyers, C. V., & Hambrick Hitt, D. (2017). School Turnaround Principals: What Does Initial Research Literature Suggest They Are Doing to Be Successful? *Journal of Education for Students Placed at Risk (JESPAR)*, *22*(1), 38–56. https://doi.org/10.1080/10824669.2016.1242070

Murphy, J. F., & Bleiberg, J. F. (2019). Insights from Turnaround Failure. In J. F. Murphy & J. F. Bleiberg (Eds.), *School Turnaround Policies and Practices in the US: Learning from Failed School Reform* (pp. 163–170). Springer International Publishing. https://doi.org/10.1007/978-3-030-01434-6_7

Papay, J., & Hannon, M. (2018, November 8). *The Effects of School Turnaround Strategies in Massachusetts*. 2018 APPAM Fall Research Conference: *Evidence for Action: Encouraging Innovation and Improvement*, Washington, D.C. https://appam.confex.com/appam/2018/webprogram/Paper26237.html

Peck, C., & Reitzug, U. C. (2014). School Turnaround Fever: The Paradoxes of a Historical Practice Promoted as a New Reform. *Urban Education*, *49*(1), 8–38. https://doi.org/10.1177/0042085912472511

Peurach, D., & Neumerski, C. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379–420. https://doi.org/10.1007/s10833-015-9259-z

Pham, L. D., Henry, G. T., Zimmer, R., & Kho, A. (2018). *School Turnaround over Time: An Extended Evaluation of Tennessee's Achievement School District and Local Innovation Zones*. annual conference of the Association for Education Finance and Policy, Portland, OR.

Public Impact. (2008). *School Turnaround Leaders: Competencies for Success*. Author.

Public Impact. (2009). *School Turnarounds: A Review of the Cross-Sector Evidence on Dramatic Organizational Improvement*. Center on Innovation and Improvement.

Rhim, L. M., & Redding, S. (2014). *The State Role in School Turnaround: Emerging Best Practices*. Center on School Turnaround at WestEd. https://eric.ed.gov/?id=ED559613

Steiner, L., & Hassel, E. A. (2011). *Using Competencies to Improve School Turnaround Principal Success*. Public Impact and University of Virginia's Darden/Curry Partnership for Leaders in Education. https://eric.ed.gov/?id=ED539544

Stuit, D. A. (2010). *Are Bad Schools Immortal? The Scarcity of Turnarounds and Shutdowns in Both Charter and District Sectors*. Thomas B. http://eric.ed.gov/?id=ED513905

Tekleselassie, A. A., & Villarreal III, P. (2011). Career Mobility and Departure Intentions among School Principals in the United States: Incentives and Disincentives. *Leadership and Policy in Schools*, *10*(3), 251–293. https://doi.org/10.1080/15700763.2011.585536

Trujillo, T., & Renee, M. (2012). *Democratic School Turnarounds: Pursuing Equity and Learning from Evidence*. National Education Policy Center. http://eric.ed.gov/?id=ED536004

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational Leadership and Student Achievement: The Elusive Search for an Association. *Educational Administration Quarterly*, *39*(3), 398–425. https://doi.org/10.1177/0013161X03253411

Zimmer, R., Henry, G. T., & Kho, A. (2017). The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones. *Educational Evaluation and Policy Analysis*, *39*(4), 670–696. https://doi.org/10.3102/0162373717705729

Appendix

Table A-3-1. Turnaround and effective principal effects on math scores

	LP as bottom 10%		LP as bottom 25%	
	(1)	(2)	(3)	(4)
Turnaround principal	0.018	0.017	-0.007	-0.008
	(0.022)	(0.022)	(0.016)	(0.016)
Effective principal	0.003	0.002	0.004	0.003
	(0.008)	(0.008)	(0.009)	(0.009)
LPS	-0.005	-0.005	-0.006	-0.006
	(0.010)	(0.010)	(0.007)	(0.007)
Turnaround principal X LPS	-0.001	-0.001	0.033	0.033
	(0.042)	(0.042)	(0.024)	(0.024)
Effective principal X LPS	-0.001	-0.001	-0.014	-0.013
	(0.022)	(0.022)	(0.018)	(0.018)
TP missing	-0.007	-0.007	-0.008	-0.008
	(0.005)	(0.005)	(0.005)	(0.005)
Principal turnover		-0.003		-0.003
		(0.002)		(0.002)
Constant	0.115^{*}	0.113^{*}	0.115^{*}	0.113^{*}
	(0.046)	(0.046)	(0.046)	(0.046)
F (TP=EP)	0.11	0.11	1.99	1.90
p-value (EP=TP)	0.736	0.743	0.159	0.168
R^2	0.716	0.716	0.716	0.716
Obs	4554232	4554232	4554232	4554232

All models include school, principal, and student covariates, and school, year, and subject fixed effects. Standard errors clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-3-2. Turnaround and effective principal effects on reading scores

	LP as bottom 10%		LP as bottom 25%	
	(1)	(2)	(3)	(4)
Turnaround principal	0.015	0.014	0.010	0.009
	(0.012)	(0.012)	(0.009)	(0.009)
Effective principal	0.009*	0.008*	0.007	0.007
	(0.004)	(0.004)	(0.004)	(0.004)
LPS	-0.003	-0.003	-0.010*	-0.011*
	(0.006)	(0.006)	(0.004)	(0.004)
Turnaround principal X LPS	-0.012	-0.012	0.006	0.005
	(0.026)	(0.026)	(0.013)	(0.013)
Effective principal X LPS	0.008	0.008	0.005	0.005
	(0.013)	(0.013)	(0.011)	(0.011)
TP missing	-0.006*	-0.006*	-0.006*	-0.006*
	(0.003)	(0.003)	(0.003)	(0.003)
Principal turnover		-0.004*		-0.004*
		(0.002)		(0.002)
Constant	0.202***	0.199***	0.201***	0.198***
	(0.026)	(0.026)	(0.026)	(0.026)
F (TP=EP)	0.23	0.25	0.06	0.03
p-value (EP=TP)	0.634	0.616	0.799	0.858
R^2	0.695	0.695	0.695	0.695
Obs	4526793	4526793	4526793	4526793

All models include school, principal, and student covariates, and school, year, and subject fixed effects. Standard errors clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-3-3. Balance checks on turnaround and other effective principals in receiving schools

Panel A. Low performing as bottom 10%

		Turnaround Ps	Other effective Ps	Diff (TP-EP)	T [p]
NCEES Standard 1 deviation		-0.189	-0.103	-0.086	-0.76
	N	32	377	409	[0.449]
NCEES Standard 2 deviation		-0.080	-0.120	0.040	0.37
	N	32	377	409	[0.713]
NCEES Standard 3 deviation		-0.120	-0.102	-0.019	-0.17
	N	32	377	409	[0.868]
NCEES Standard 4 deviation		-0.080	-0.071	-0.009	-0.08
	N	32	377	409	[0.937]
NCEES Standard 5 deviation		-0.096	-0.065	-0.031	-0.29
	N	32	377	409	[0.773]
NCEES Standard 6 deviation		-0.188	-0.044	-0.143	-1.34
	N	32	377	409	[0.180]
NCEES Standard 7 deviation		-0.074	-0.143	0.069	0.59
	N	32	377	409	[0.557]
NCEES median deviation		-0.126	-0.082	-0.044	-0.42
	N	32	377	409	[0.674]
School leadership deviation		-0.111	-0.006	-0.105	-1.19
	N	17	256	273	[0.236]
Teacher leadership deviation		-0.110	-0.023	-0.087	-0.94
	N	17	256	273	[0.347]
Community support deviation		-0.167	-0.034	-0.132	-1.64
	N	17	256	273	[0.102]
Managing student conduct deviation		-0.081	-0.026	-0.055	-0.57
	N	17	256	273	[0.571]
Teacher turnover rate deviation		0.034	0.020	0.014	0.77
	N	38	446	484	[0.443]
Observations		488			

Panel B. Low performing as bottom 25%

		Turnaround Ps	Other effective Ps	Diff (TP-EP)	T [p]
NCEES Standard 1 deviation		-0.098	-0.113	0.015	0.21
	N	90	319	409	[0.838]
NCEES Standard 2 deviation		-0.071	-0.129	0.058	0.82
	N	90	319	409	[0.410]
NCEES Standard 3 deviation		-0.079	-0.110	0.031	0.43
	N	90	319	409	[0.668]
NCEES Standard 4 deviation		-0.019	-0.087	0.067	0.96
	N	90	319	409	[0.340]
NCEES Standard 5 deviation		-0.042	-0.075	0.032	0.46
	N	90	319	409	[0.643]
NCEES Standard 6 deviation		-0.045	-0.059	0.014	0.19
	N	90	319	409	[0.846]
NCEES Standard 7 deviation		-0.131	-0.140	0.009	0.12
	N	90	319	409	[0.907]
NCEES median deviation		-0.041	-0.098	0.057	0.83
	N	90	319	409	[0.406]
School leadership deviation		-0.106	0.011	-0.117^{*}	-2.23
	N	56	217	273	[0.027]
Teacher leadership deviation		-0.114	-0.006	-0.108	-1.97
	N	56	217	273	[0.050]
Community support deviation		-0.152	-0.014	-0.138^{**}	-2.88
	N	56	217	273	[0.004]
Managing student conduct deviation		-0.078	-0.016	-0.062	-1.07
	N	56	217	273	[0.286]
Teacher turnover rate deviation		0.020	0.021	-0.001	-0.12
	N	103	381	484	[0.902]
Observations		488			

Sample restricted to principals classified as turnaround or effective principals in their first year in a receiving school. Estimates from t-tests comparing deviations from school-level means of row variables in principal's prior school.