

PRIMING THE LEADERSHIP PIPELINE:
SCHOOL PERFORMANCE AND CLIMATE UNDER AN URBAN SCHOOL LEADERSHIP
RESIDENCY PROGRAM

Jennifer L. Steele, Elizabeth D. Steiner, Laura S. Hamilton
2020

Abstract

Purpose. This study examines school climate and student achievement trends under an ambitious school leadership residency program in an urban school district. The two-year leadership residencies were intensive, combining at least 370 hours of professional development with on-the-job training, in which aspiring school principals held either assistant-level administrative or teacher leadership roles.

Research Design. Using a difference-in-differences framework with school fixed-effects, we estimate the relationship between schools' cumulative exposure to program residents and measures of school climate and student performance. We measure school climate using school-by-semester teacher survey composites. Student performance is captured using school-by-year data on language arts and math scale scores, chronic absence rates, suspension rates, and graduation rates.

Findings. In models that allow average time trends to vary between the state and the treatment city, an additional resident-by-year in an administrative role in high schools is linked to an additional 15% of a school-level standard deviation in math scale scores and an additional 3.6 percentage points in graduation rates, but also to an additional 10 percentage points in suspension rates. Results are sensitive to model specification, school level, and to residents' placement in administrative or teacher leader roles.

Implications. Due to the contracting nature of the district, only one of 30 entering residents became a school principal within three years of program inception. In some models, the

estimates suggest potential for aspiring leaders to effect change from non-principal administrative roles. Potential for teacher leadership roles is less clear.

Keywords: urban education, principal preparation, instructional leadership, teacher leadership, school climate, student achievement

Article Type: Empirical paper

Introduction

The question of how to improve struggling schools lies at the heart of national concerns about public education quality. Policy responses to that question often emphasize the cultivation of strong school leadership. Under the 2001 No Child Left Behind Act, schools that failed to make Adequate Yearly Progress over five consecutive years were required to choose among several restructuring options, all of which entailed the appointment of new leadership (Hassel, Hassel, Arkin, Kowal, & Steiner, 2006). A 2008 practice guide by the U.S. Department of Education's What Works Clearinghouse identified the signaling of change through strong leadership as its first of four evidence-based recommendations for turning around low-performing schools (Herman et al., 2008).

Empirical evidence suggests that principals exert measurable influences on school effectiveness as measured by standardized test scores. In a research synthesis, Leithwood, Louis, Anderson, & Wahlstrom (2004), argued that school leadership explained 3% to 5% of variation in student learning between schools, or about a quarter of the variance attributable to school-based variables. More recent studies have attempted to directly estimate the magnitude of school leader effects using administrative datasets. Depending on the study, the percent of between-principal variance in single-year school effects has been estimated as about 6% of graduation rates (Coelli & Green, 2012), 7% of reading and math achievement (Chiang, Lipscomb, & Gill,

2016), 3-6% of reading and 6-8% of math achievement (Grissom, Kalogrides, & Loeb, 2015), and 10% of reading and 14% of math achievement (Dhuey & Smith, 2014, annualized linearly from cumulative estimates). A broad lesson is that the variation in principal effects may be nearly as large as teacher effects (Aaronson, Barrow, & Sander, 2007; Branch, Hanushek, & Rivkin, 2012; Kane & Staiger, 2005; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004).

Some evidence suggests that a portion of principals' effects may be exerted through their influence on school climate. We follow Cohen, McCabe, Michelli, & Pickeral (2009) and the National School Climate Council (2007) in defining school climate as the "quality and character of school life," reflecting "norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures" (Cohen et al., 2009, p. 182). For instance, Goddard, Salloum, & Berebitsky (2009) found within a stratified random sample of 78 Michigan elementary schools that teachers' reported sense of trust in their school communities mediated relationships between student disadvantage and academic achievement, implying that negative links between student disadvantage and achievement operate through detriments to trust. In a separate study, Grissom (2011) used nationally representative data to find that principals whose leadership teachers rated highly also faced much lower rates of teacher turnover, especially in disadvantaged schools. This finding is notable in light of evidence that higher teacher turnover may harm student achievement on average (Ronfeldt, Loeb, & Wyckoff, 2013).

Given that principal leadership matters, the question of how school systems should cultivate and develop strong leaders presents an important challenge. Of special policy concern is the question of how to prepare leaders with the skills to improve student outcomes in schools with low historic levels of achievement.

The current article provides results from a leadership development partnership among a small urban school district, three local charter management organizations (CMOs), and a nonprofit organization, TNTP (formerly, The New Teacher Project), which helps schools and districts build human resource capacity. The partnership was called Pathways to Leadership in Urban Schools (PLUS). Funded by a federal School Leadership Program grant, PLUS was designed to create a pipeline of well-prepared school leaders who would generate systematic improvement for the city's schools over time. These leaders would be drawn from local schools and would receive intensive, hands-on training on improving academic instruction and school climate. Our analysis focuses on the first three years of the program's implementation, the school years 2014-15 through 2016-17. During these years, and through 2017-18, potential school leaders were selectively admitted to a two-year residency program, in which they worked as assistant-level school administrators or as classroom teachers while receiving intensive coaching and professional development around instructional leadership and school climate. After successfully completing the residency program, those who did not already hold administrative licensure in the state were eligible to receive it.

Our study examines trends in students' math and English Language Arts (ELA) proficiency associated with staffing by PLUS residents, as well as trends in schools' chronic absence and attendance rates, their out-of-school suspension rates, and (for high schools) their four-year graduation rates. It also examines resident-associated trends in teachers' reports of the professional climates of their schools. Using school fixed-effect models that also allow average time trends to vary between the state and the treatment city, an additional resident-by-year *working in an administrative role in high schools* is linked to an additional 15% of a school-level standard deviation in math scale scores and an additional 3.6 percentage points in graduation

rates, but also to an additional 10 percentage points in high school suspension rates. An additional resident-by-year working *in a teacher leadership role* is linked to 0.16 of a school-level standard deviation lower ELA test scores in elementary and middle schools, but to 0.23 of a school-level standard deviation higher math performance in high schools. Results are sensitive to model specification, school level, and to residents' placement in administrative or teacher leader roles. Schools' exposure to PLUS residents appears largely unrelated to teachers' reports of the educational climates of their schools.

This article is organized as follows: The next section summarizes related research on school administrative training and professional development programs. We then describe the PLUS program, its geographic context, and its underlying theory of action. Next, we describe our data sources and analytic methods. This is followed by our results in terms of the relationship of residency placement and dosage to schools' culture, academic performance, behavioral outcomes, and graduation rates. We conclude with a discussion of limitations and implications.

Literature on School Leadership Preparation and Training

Only a few studies have attempted to quantify the quality of school leadership preparation programs on a large scale. In a national study commissioned by The Wallace Foundation, Darling-Hammond, LaPointe, Meyerson, Orr, and Cohen (2007) found that principal preparation programs rated most highly by their graduates were those that rigorously selected applicants, emphasized instructional leadership, prioritized practical skills, recruited educators familiar with the needs of the local community, and promoted hands-on learning experiences. More recently, Grissom, Mitani, and Woo (2018), tracked graduates from twelve university-based principal preparation programs in Tennessee. They found that the programs ranked differently depending on the outcomes under consideration, such as supervisor ratings,

teacher ratings, and student achievement growth, and on the types of schools into which program graduates were hired.

Other studies that have estimated the effects of principal preparation on student achievement have focused on programs offering alternatives to university-based licensure programs. In most U.S. states, school administrative licensure requires several years of classroom teaching experience, passing a licensure test, and obtaining a graduate-level degree in school administration or leadership (Briggs, Cheney, Davis, & Moll, 2013; Campbell & Gross, 2012). Proponents of alternative routes have argued that such requirements deter talented candidates and provide insufficient practical preparation in how to raise student achievement (Campbell & Gross, 2012; Levine, 2005). Alternative licensure programs that include on-the-job training and embedded professional development have arisen to address regional shortages in licensed school leaders as well as concerns about variation in preparation quality (Herrington & Wills, 2005).

Corcoran, Schwartz, and Weinstein (2009, 2012) examined three-year effects of principals trained by New York City's Aspiring Principals Program, a selective, alternative-route principal preparation program that allows new school leaders to bypass university licensure and assistant principal roles. The program focused on the improvement of student achievement as well as on organizational management skills. The study found modest positive effects of program-trained principals on students' achievement trajectories in some models, particularly in ELA, but results were highly sensitive to model specification.

Two studies have focused on the effects of principals prepared by New Leaders for New Schools, an alternative-route program emphasizing instructional leadership as well as organizational management skills. In a study that examined the effects of New Leaders-trained

principals on student achievement in Oakland Unified School District, Booker and Thomas (2014) found that students in schools that had had New Leaders principals for at least three years outperformed others in Oakland by an average of 6% of a student-level standard deviation in ELA, and 16% of a standard deviation in math. In a ten-district evaluation of the performance of New Leaders principals, Gates, Hamilton, Martorell, Burkhauser, and colleagues (2014) showed that students in schools staffed by a New Leaders principal with at least three years of experience outperformed other schools in the district, controlling for principal experience and observed school and student characteristics, by about 2.5% of a student-level standard deviation in reading and mathematics in elementary grades, though results varied markedly by district. The researchers also examined whether leadership effects were attenuated in high schools due to their greater size, and whether they were intensified in charter schools due to alignment with the school reform mission of the New Leaders program. The estimated effect did indeed differ in high schools, with about 7.5% of a standard deviation benefit in reading and no effect in mathematics. But effects did not differ between New Leaders principals placed in traditional and charter schools. In a follow-up report, the researchers noted that because New Leaders benefits became evident only after program completers had been hired as and served at least two to three years as principals, pipeline programs can take many years to show effects (Gates, Baird, Doss, et al., 2019). They also noted that within-district comparisons like the ones they conducted may understate effects of program-trained leaders if districts are adopting new leadership development approaches district-wide.

In a Wallace Foundation-funded study of six districts' efforts to create their own principal pipelines, Gates, Baird, Master, and Chavez-Herrerias (2019) identified four components of district pipeline strategies: developing or revising standards for leadership,

creating new systems of pre-service preparation, selectively hiring and placing leaders, and providing on-the-job leadership support and evaluation. After three years on the job, principals newly hired in districts undertaking such pipeline reforms outperformed newly hired principals in similar districts by 6.22 percentile points in students' reading achievement, and by 2.87 percentile points in students' math achievement. Though estimates varied by district, the effects of particular components of the pipeline reforms could not be identified.

Adding to the research about school leader support and training, two recent randomized field trials have examined the effects of professional development programs on the performance of in-service principals. Jacob, Goddard, Kim, Miller, & Goddard (2015) randomized 126 schools in rural Michigan to the McREL Balanced Leadership program, which provided ten two-day professional development sessions to treatment-group principals, focused largely on instructional leadership. Researchers hypothesized relationships among principal leadership, instructional climate (defined as teacher trust and collective efficacy as well as norms for collaboration and differentiated instruction), teacher turnover, and student achievement. In practice, the three-year study found no effect on student achievement in math or reading or on teacher surveys about the instructional climates of their schools, but treatment schools were found to have lower subsequent teacher turnover. Fryer (2017) conducted a randomized trial of an intensive professional development program for in-service Texas principals focused on instructional leadership. The intervention included 170 hours of principal training workshops per year for two years. It found positive first-year effects on students' math and reading test scores of about 8% of a student-level standard deviation, but no statistically significant effects in the second year.

Research on the effects of teacher leadership in schools, though less definitive, is also encouraging. Descriptive evidence suggests that a school climate marked by frequent instructional collaboration among teachers is associated with higher student achievement (Goddard, Goddard, & Tschannen-Moran, 2007). A meta-analysis of causal evidence has shown that instructional coaching of teachers can produce substantial positive effects on student achievement (Kraft, Blazar, & Hogan, 2018), but the study did not examine which studies used other classroom teachers as coaches. Papay, Taylor, Tyler, and Laski (2016) conducted a field experiment in which low-performing teachers received instructional coaching from higher-performing peer teachers. They found positive effects of 0.12 of a student-level standard deviation on the effectiveness of the teachers receiving coaching. Effect estimates were modestly larger in the year after treatment ended, suggesting that effects persisted and may have been cumulative. Their results were consistent with those of Jackson and Bruegmann (2009), who found student-level effects of 0.03 to 0.04 of a standard deviation when the effectiveness of teachers' peers in the same subject and grade rose by a standard deviation. These findings are notable because they suggest that teachers may influence their peers' instruction through both formal and informal avenues.

Research Questions

Building on this body of research, our study examines changes in school climate, achievement, behavior, and attainment associated with schools' exposure to educators who participated in or graduated from the PLUS residency program. Ours is the first study we know of to examine results linked to leaders in alternative-route licensure programs working in subordinate leadership positions as assistant administrators or as teachers with leadership

responsibilities. Leveraging within-school variation over time in schools' exposure to aspiring principals, we address two research questions:

1. How is the placement of PLUS residents in a school related to teachers' reports of school climate? And do these relationships differ in high schools versus elementary and middle schools, or in charter versus traditional schools?
2. How is the placement of PLUS residents in a school related to students' academic, behavioral, and attainment outcomes of interest? Do these relationships differ in high schools versus elementary and middle schools, or in charter versus traditional schools?

Our interests include differential effects for high schools and charter schools in light of the aforementioned research showing that principals prepared through alternative routes may exert different effects on high schools (Gates et al., 2014) and that charter schools may offer working environments that are better aligned with the reform focus of alternative leader preparation programs (National Alliance for Public Charter Schools, 2008). The fact that 19% of charter school leaders also teach, as compared to 2% of public school leaders, suggests that charter schools may be particularly receptive to roles that combine features of school leadership and teaching (National Alliance for Public Charter Schools, 2008).

Leadership Residency Context and Theory of Action

This article examines outcomes associated with the first three years of the PLUS program's implementation in a small U.S. city. Due to a decline in its manufacturing base, the city's population size contracted by about 40%, to fewer than 100,000 people, over the past seventy years. In the past 15 years, the city has faced poverty rates that were among the highest in the United States. Between 2003 and 2016, the average percentage of students qualifying for free and reduced-price lunches in the city's public schools was about 90%, as compared to 36%

for the remainder of the state. During the same years, schools' academic proficiency on state accountability tests lagged the rest of the state by about 18 to 30 percentage points.

In 2013, the city's public school district partnered with TNTP to develop a leadership residency program that would prepare a new cadre of school principals for the city, and the partnership received a five-year School Leadership Program grant from the U.S. Department of Education. PLUS recruited its first cohort of school leadership residents in the spring of 2014 and launched its first five-week summer institute that summer. Our study examines this first cohort and two subsequent cohorts that were recruited in spring 2015 and 2016, respectively. The PLUS program assisted residents with finding full-time employment as assistant administrators (or in subsequent years, as either administrators or teachers), in district-run or charter schools in the city.

<Insert Figure 1 about here>

A logic model outlining PLUS's theory of action is shown in Figure 1, and additional details are provided in ensuing subsections. The program's *inputs* to the local schools included the careful *selection* of promising leaders, largely drawn from educators already working in and familiar with the community; an intensive *coaching and professional development* process designed to equip school leaders with the instructional leadership and human capital management skills needed to cultivate strong learning environments; and an *on-the-job learning* experience in which insights from the coaching and professional development could be immediately applied. These inputs were expected to yield progress in the near-term (program years one to three) as well as the medium-term (program years three to five), with the idea that having a strong cadre of leaders would eventually lead to long-term district transformation. Specifically, the program anticipated that the instructional leadership work undertaken by the

residents would begin to improve school *processes*, including instructional quality as well as other school climate concepts measured on TNTP's semi-annual teacher school climate survey, such as teachers' collaborative planning and use of data, their access to quality instructional feedback, their sense of being fairly evaluated, and their sense that their school leaders provided a safe and orderly learning environment. The program anticipated that these changes would, in turn, produce evidence of improved *outcomes* such as stronger academic achievement, higher attendance rates, lower suspension rates, and higher graduation rates. The model specified that these processes and effects could begin to emerge in the near-term as residents pursued leadership expertise in their schools, but that they were likely to intensify in the medium term as residents moved into principalships where they would have authority to set the priorities for their schools.

Informed by more than a decade of research showing annual principal turnover rates ranging from 15% to 30% (Béteille, Kalogrides, & Loeb, 2012), the launch of the program was undergirded by the notion that some residents could begin filling principalships as soon as their second residency year. But in light of the city's shrinking population, the number of public schools in the city declined during the treatment period. The 2015-16 and 2016-17 academic years saw the openings of one to two principalships per year, and these were not guaranteed for PLUS residents or alumni. As of the 2016-17 school year, one resident, a member of the entering 2015-16 cohort, had been hired into a principalship.¹

¹ Two recent large-scale studies have cautioned that a lag between administrative licensure and the principalship is to be expected. Using North Carolina administrative data, Bastian and Henry (2015) found an average wait time of 5.12 years among those who did eventually become principals. Using administrative Texas data and tracking teachers from administrative licensure forward, Davis, Gooden, & Bowers (2017) found that only 20% became principals within six years, and that fewer than half became principals within the 16 years observed in the data. Also, after tracking graduates from twelve principal preparation programs in Tennessee, Grissom, Mitani, and Woo (2018) found that between 28% and 52% were hired as assistant principals within five years, and that only 6% to 17% were hired as principals within five years.

By the second implementation year (2015-16), the leadership residency program had also become part of the district's partnership with a local CMO that had agreed to assume operation of several of the district's lower-performing schools. This CMO became part of the school leadership residency program by providing administrative slots for several PLUS residents.

Resident Selection

PLUS residents were selected through a rubric-scored, three-phase application process: an online application with essay questions and a lesson evaluation task; a 30-minute phone interview about leadership experiences and goals; and a half-day meeting involving interviews, group work, and role-plays. The program prioritized residents who not only had teaching experience but presented evidence of having raised student achievement, had taught in the treatment city or demonstrated a commitment to the city, had demonstrated leadership potential in their schools, and were recommended by respected peers in the school system.

In the first year, 134 individuals applied to be part of the inaugural cohort. Of the 16 eventually selected, nine found residency employment in the city, yielding a 9.7% acceptance-and-placement rate among those who applied. In subsequent years, the ratio of initial applicants to accepted residents fluctuated only moderately around the initial figures.

Program Components

Each new cohort began the program by taking part in the PLUS summer institute, which ran seven to nine hours per day for five weeks in the summer prior to the start of school. The institute sessions, which included group work, simulations, and role-plays, focused on topics such as observing and evaluating teaching lessons, providing "bite-size" and actionable feedback to teachers, and working effectively with mentor principals.

During their first residency year, PLUS residents met bi-weekly at their respective school sites with a PLUS leadership coach. Through these one-on-one meetings, the coaches helped residents address specific challenges, with emphasis on time management, obtaining supervisors' and teachers' support for instructional coaching, and balancing residency duties (namely, observing and coaching teachers) with other responsibilities of their residency jobs, such as handling non-instructional administrative tasks or teaching classes.

Residents in their first two years also participated in monthly, daylong professional development workshops led by PLUS staff and coaches. They also took part in virtual coaching, in which they videotaped their instructional coaching sessions and discussed them with an online leadership coach, and they completed monthly assignments that culminated in an instructional improvement plan for their respective schools. Altogether, the program provided at least 300 hours of professional development to each resident during the first residency year and about 70 hours during the second year, bringing the two-year total to at least 370 hours per resident. Residents who successfully completed the program received an administrative license endorsement if they were not already licensed.

Residency Placements

Residents were placed into either administrative or teacher leadership roles. Among the first cohort of residents, all applicants already held state licensure, but this was not the case for the subsequent two cohorts. Table 1 indicates the distribution of initial residency placements for members of each of the first three cohorts, as well as the number who completed the first residency year in each cohort, overall and by initial residency role.

<Insert Table 1 here>

Advent of Administrative Roles

Eight of nine members of the first cohort found employment in *formal administrative roles* called *Lead Educator*, in which their responsibilities included instructional leadership, test coordination, and tasks comparable to those of an assistant principal. One of the nine was hired in a charter school as an *Apprentice School Leader (ASL)*. ASL roles were also administrative. They were specific to charter schools and typically emphasized teacher observation and instructional coaching. Both Lead Educator and ASL roles were similar to the roles of an assistant principal.

Creation of Teacher Leader Roles

In Cohort 2, two residents were hired into Lead Educator roles in district schools, and four obtained ASL roles in charter schools, meaning that six residents were placed as administrators. However, Cohort 2 also saw the introduction of a new role: *Teacher Leader*, into which six cohort members who lacked initial licensure were placed. Teacher Leaders were employed in district schools as full-time classroom teachers but, as part of the residency program, were expected to observe and coach a small caseload of fellow teachers on their instruction.

In Cohort 3, six residents obtained positions as Teacher Leaders in district schools (two already licensed), and two others were hired as administrative ASLs in charter schools. Scope and support for the Teacher Leader roles were defined largely by the hiring schools. A few Teacher Leaders reported that they received a release period for such coaching, but most said they did not.

Data Sources and Sample

To examine the relationship between schools' exposure to PLUS residents and teachers' reports about the professional culture in their schools, we employ semi-annual teacher survey

data collected by the residency program from teachers across the city from 2013-14 through 2016-17. We use school-by-year data from the state for the school years 2005-2006 through 2016-2017 to estimate the relationship between placement/dosage of PLUS residents in the city's schools and the schools' achievement, behavioral, and attainment outcomes over time. In addition, we gathered qualitative data on participants' experiences in the program through annual focus groups and interviews with PLUS residents. We present qualitative findings, including recommendations for how principals might support residents' roles—especially the Teacher Leader role—more consistently, in a separate paper focused on participants' experiences (Steele, Steiner, & Hamilton, 2018).

Treatment Variables

Our independent variables of interest capture schools' exposure to PLUS residents and alumni (described henceforth, collectively, as "residents"). Our simplest treatment measure is a dichotomous time-varying school-level variable, *has resident*, which is coded 1 if the school is staffed by one or more PLUS residents in a given year, and 0 otherwise. But in light of evidence that effects of both school leadership and peer coaching may be persistent and cumulative (Dhuey & Smith, 2014; Papay et al., 2016), we are also interested in the effects of a school's cumulative intensity of exposure to PLUS residents. Leadership effects may emerge gradually as leaders have time to enact new school policies, hire and develop teachers, and counsel out ineffective teachers (Boyd et al., 2011; Gates et al., 2014; Leithwood et al., 2004)—behaviors that some PLUS residents described enacting (Steele, Steiner, & Hamilton, 2018)—and because schools may benefit from a critical mass of leadership residents who share the same training and perspectives. Indeed, PLUS residents described perceiving more support for their work in schools that had employed more PLUS residents (Steele, Steiner, & Hamilton, 2018). We first

measure treatment dosage as the *cumulative number of residents-by-year* who have staffed a given school as of a given time period. Because we are also interested in whether residents' effects depended on whether they were placed in an administrative or teacher leadership role, we create two additional dosage variables. *Cumulative administrative dosage* is defined as the cumulative number of residents-by-year who have staffed the school in administrative roles as of a given time period, whereas *cumulative teacher leadership dosage* is the cumulative number of residents-by-year who have staffed the school in teacher leadership roles at a given point in time. For example, if the school had one resident in an administrative role in 2014-15, two residents in administrative roles in 2015-16, and one resident in an administrative role in 2016-17, its value of cumulative administrative dosage would be one in 2014-15, three in 2015-16, and four in 2017-18.²

School Climate Measures

With respect to the logic model, we measure the "process" variables related to school climate using eight waves of data from a teacher survey developed and administered by PLUS's parent organization, TNTP. The 40-item survey was administered across the district and in partner CMOs in winter and spring of each school year from the year prior to PLUS' launch, 2013-14, through 2016-17. The survey, which reportedly required 15 to 20 minutes for teachers to complete, was administered to all teachers in a given school at each time point, and principals were provided with a school-level report after each survey administration. The survey was developed and validated by TNTP and is proprietary.

² When we define dosage as number of residents in a school in a given year—i.e., as concurrent rather than cumulative—estimates tend to be very slightly larger in the magnitudes of the absolute values, but are otherwise substantively and statistically nearly identical to those reported here. They are available upon request.

For this study, we had access to teacher-by-item level data for each survey administration in the city through the 2016-17 school year. School identifiers were stable, but anonymized teacher identifiers changed each year. We construct five composites from the items based on logical coherence and parsimony: *learning environment and leadership*; *peer collaboration for student growth*; *observation, feedback, and professional development*; *fair evaluation*, and an *overall school climate* composite that is an aggregate of the other composites. The composites, which reflected leadership priorities emphasized in PLUS professional development, are described in Table 2. The overall composite is based on 34 items; the other composites are means of between 7 and 10 items each. The underlying items range from 0 to 10, with 10 being the highest level of agreement, and 0 being the lowest. As shown in Table 2, the composite means were close to 7, and standard deviations were approximately 2. Their reliability estimates ranged from a high of 0.95 for the overall composite to a low of 0.9 for the measure of peer collaboration and focus on student growth. In addition, we examine a standalone item, *years plan to stay*, which is the number of years beyond the current one that the teacher said he or she planned to stay in the current school. It ranges from 0 to 10+ in units of 1 (rescaled as 0 to 10), with a mean of 6.4 and a standard deviation of 3.8.

<Insert Table 2 about here>

Schools' Academic, Behavioral, and Attainment Outcomes

To capture students' academic and behavioral outcomes in treatment and comparison schools, we use school-by-year data published by the state for the school years 2005-2006 through 2016-2017. We focus on school-level measures because the treatment indicators-- placement and dosage of leadership residents--were designed as school-level interventions, and information was not available on which teachers received coaching by residents. Our dependent

variables in terms of achievement are *school-level standardized scale scores* in math and ELA. These scores are reported by the state at the school-by-grade-by-year level, including subject-area math scores (i.e., algebra and geometry) for the high school grades. To create comparability of scale scores between grades and between years, we standardize them to have mean 0 and standard deviation 1 by grade level and year, using the standard deviation of school-level scores in the state within grade and year.³ We then weight these standardized scores by number of students in each grade in a given year in the school, and we average the weighted, standardized scores to the school-by-year level for the spring of 2006 through 2017.

Our analysis also examines school-level measures of student behavior, including *chronic absence rates*, which are available in the state's school accountability reports for 2012-13 through 2016-17, and *attendance rates*, which are available only for the treatment district from 2013-14 through 2016-17. A school's percent chronically absent is the share of enrolled students absent more than 10 percent of school days during the year. Its attendance rate is the average percent of students in attendance at least part of the school day over the course of the school year. We also examine *suspension rates*, which are available via the state's school accountability reports for 2004-05 through 2016-17. A school's suspension rate is the percentage of enrolled students who received at least one out-of-school suspension during the academic year.

For high schools, we examine schools' academic performance based on their *four-year graduation rates*, which are the percentages of ninth graders, adjusted for transfers in and out of the school, who earn high school diplomas from the school within four years. This measure is policy-relevant given the importance of high school completion for subsequent educational and

³ The state changed from its longstanding accountability test to the Partnership for Assessment of Readiness for College and Careers (PARCC) test of the Common Core State Standards in 2014-15.

economic opportunity (Cameron & Heckman, 1993; Clark & Martorell, 2014; Psacharopoulos & Patrinos, 2018), and given the average four-year completion rate of 55% in ever-treated schools in the treatment district in 2014-15, the year the residency program was launched.

Table 3 displays descriptive statistics for outcome and control variables in schools within and outside of the treatment city, disaggregated by *ever-treated* versus *never-treated* status, meaning whether the school had ever been staffed by PLUS resident as of 2016-17. The statistics presented refer to 2014-15, the launch year of PLUS program placements. As expected, the ever-treated schools markedly underperformed the rest of the state in academic outcomes; they also showed double the level of chronic absence and four times the rate of suspensions. But they also modestly underperformed the rest of the district, with lower ELA and math scores and proficiency rates, as well as lower graduation and higher suspension rates, though their attendance rates were similar.

<Insert Table 3 about here>

Control Variables

We capture a variety of school-by-year characteristics from state accountability report cards. These include school level indicators for elementary, middle, and high schools; a dichotomous charter school indicator; a measure of the total enrollment; a measure of student/teacher ratio; the percentages of students identified as Asian, Black, and Hispanic, respectively; and the percentage who qualified for free or reduced-price lunch. In some model specifications, we also control for ever-treated status, which is, again, a time-invariant indicator of whether a school had ever been staffed by a resident as of 2016-17. Thirteen schools in the dataset were ever-treated, whereas 24 schools in the treatment city were never-treated, including

a handful of charter schools that were not explicitly part of the treatment partnership. Statewide, 2,465 schools in the dataset were never-treated.

These control variables are summarized in Table 3. Never-treated and ever-treated school within the treatment city were similar, though high schools were overrepresented in the ever-treated sample, at 23% versus 8%. In comparison to the rest of the state, ever-treated schools were much more likely to be charter (39% versus 4%) and had much higher concentrations of students who were Black, Hispanic, and eligible for subsidized meals.

Analytic Strategy

Our analysis is designed to illuminate the relationship between PLUS residency placements and a variety of school-level achievement, behavioral, and attainment outcomes over time, as well as several school climate attributes as reported by teachers. Our preferred specification uses school fixed effects to eliminate selection bias resulting from unobserved differences in which schools received residents. The model is specified as in Equation 1:

$$y_{st} = \alpha + \boldsymbol{\phi}'\mathbf{year}_t + \eta tcity_s + \boldsymbol{\gamma}'(\mathbf{year} * \mathbf{tcity})_{st} + \boldsymbol{\delta}'\mathbf{X}_{st} + \boldsymbol{\theta}'\mathbf{S}_s + \beta trt_{st} + \varepsilon_{st} \quad (1)$$

where y_{st} is a school-level outcome for school s in year t . These outcomes are predicted as a function of \mathbf{year}_t , which is a vector of year dummy variables used to capture secular trends from spring 2006 through 2017, and \mathbf{x}_{st} , which is a vector of time-varying school-by-year controls that may plausibly be correlated with both treatment status and outcomes, including total enrollment size, percent subsidized-meal eligible, racial/ethnic percentages, student-teacher ratios, and charter status. For student performance outcomes in which data are available for all schools in the state, we fit some specifications in which the year dummy variables are interacted with a dichotomous treatment-city indicator ($tcity_s$) to account for vector $\boldsymbol{\gamma}$ of differential secular

trends between the treatment city and the rest of the state. The model also includes S_s , a vector of school indicators to capture time-invariant school fixed effects given by Θ . These school fixed effects allow schools to function as their own controls, removing potential confounding by stable between-school differences, and thus improving our ability to justify causal inferences.

The treatment variable of interest, defined as a time-varying *has residents* indicator or as a cumulative dosage indicator, is represented by trt_{st} , with an estimated effect given by parameter β . Note that we use the term "effect" here to characterize the statistical association between the independent and dependent variables of interest. Following Cameron & Miller (2015) and Bertrand, Duflo, and Mullainathan (2004), we cluster standard errors at the school level.

To test for differential effects in high schools versus elementary and middle schools, and for differential effects by charter versus traditional schools, we include statistical interactions between treatment and high school status or between treatment and charter status as variations of equation 1. But because these variables have little to no within-school variation over time, we replace the school fixed effects in the disaggregation models with a vector of time-invariant school characteristics, including the school grade-level category, a dichotomous ever-treated indicator, and a dichotomous indicator of whether the school is located in the treatment city, and we include a school-level random intercept in the error term to account for nesting of students within schools.

Results

School Climate Outcomes

Research question 1 asks how the placement of PLUS residents in a school is related to teachers' survey reports about various dimensions of school climate. Figure 2 presents unadjusted time-series trends in each of the school climate composites in the treatment city between 2013-

14—the year before the school leadership residencies were launched—and 2016-17. The dashed lines represent schools that had ever had a PLUS resident as of 2016-17 (n=13), and the solid lines represent schools that had not (n=24 in the survey years). Recall that school climate dimensions range from 0 to 10, with 10 indicating the strongest agreement that the dimension is present in the school, and 0 indicating the lowest agreement.

<Insert Figure 2 about here>

An important detail in Figure 2 is that the composite measures of school climate were improving across the city's public schools during the study period in both never-treated and ever-treated schools. The question relevant to this study, of course, is whether schools that employed PLUS residents made greater school climate improvements than those without residents.

To better isolate the effects of residency placement or cumulative dosage on school climate in a given year, we present regression coefficients and standard errors from school fixed-effects models in Columns 1-6 of Table 4. Panel A presents coefficients and standard errors from a dichotomous placement variable indicating the presence of one or more PLUS residents in a school in a given year. Panel B presents effects of the cumulative number of residents-by-year who have staffed a school as of a given year. Panel C focuses on the cumulative number of residents-by-year who have served in administrative roles (Lead Educator or ASL), and Panel D, on cumulative residents-by-year in Teacher Leader roles.

We find no statistically significant relationships between PLUS resident placement or dosage and any composite measure of school climate as reported by teachers.

Differential Effects by School Level and Type

In Table 5, we examine whether treatment effect estimates differed in high schools versus elementary and middle schools (Panel A) and in charter versus traditional schools (Panel B). For

parsimony, Table 5 includes only the cumulative dosage variables in the high school interaction models, and it includes only cumulative administrative dosage in the charter school interaction models, because all of the charter school residency roles were administrative (ASL) roles. In Panel A, we find some evidence that higher Teacher Leader dosage over time is linked to lower satisfaction with school climate on several dimensions in high schools, including overall climate, peer collaboration for student growth, and observation and professional development ($p < 0.05$). All three dimensions arguably pertain to the peer-coaching role Teacher Leaders were expected to play. The relationships remain indistinguishable from zero for elementary and middle schools. In Panel B, we find that higher exposure to administrative residents is linked more negatively to school climate reports in charter schools than in traditional schools ($p < 0.05$).

<Insert Table 5 about here>

School Achievement, Behavioral, and Attainment Outcomes

Our second research question concerns the relationship of schoolwide academic and behavioral outcomes to the placement of PLUS residents in a school. We first present unadjusted time trends for these outcomes in Figure 3, where the dotted lines represent ever-treated schools that were staffed by at least one PLUS resident between 2014-2015 and 2016-2017 ($n=13$), and the dashed lines represent schools in the treatment city that had never had a resident as of 2016-17 ($n=33$, or 24 in treatment years). The solid line represents never-treated schools in the rest of the state ($n=2,772$). A vertical dotted line at 2014-2015 represents the inaugural placement year of PLUS residents.

<Insert Figure 3 about here>

In the first column of Figure 3, we observe that standardized scale scores in the treatment district were improving relative to the state in both never-treated and ever-treated schools,

though they still lagged far below the rest of the state. Also, ever-treated schools trailed never-treated schools in the district by about half a standard deviation. For percent chronically absent, attendance rates, and suspension rates, we find similar trends in ever-treated and never-treated schools, with modest positive spikes in chronic absence and suspension rates in ever-treated schools during recent years. We find markedly lower graduation rates in ever-treated than in never-treated high schools, which may reflect the fact that the district's never-treated high schools were disproportionately magnet or charter schools.

Given that pre-treatment trends were not always parallel between the treatment district and the rest of the state in Figure 4, we present two sets of school fixed-effect regression estimates in Table 6. The four treatment variable specifications (treated, cumulative dosage, and cumulative dosage by role) are represented in Panels A through D. Columns 1 through 6 account for the aforementioned control variables and year dummy variables; columns 7 through 12 also include interactions of each year dummy variable with the treatment city indicator to adjust for differences in time trends between the treatment city and the rest of the state. Including this adjustment improves our ability to attribute effects to PLUS resident exposure by parceling out differential pre-treatment trends, but it also means that our secular-trend estimates are based on only 24 non-treated schools within the district, greatly limiting their precision. We present both sets of estimates in Table 6 because results are sensitive to this modeling choice.

<Insert Table 6 about here>

With respect to math scale scores, columns 1 through 6 show positive and statistically significant relationships to resident placement and dosage for all four definitions of the treatment. In Panel B, each additional resident-by-year is associated with an additional 0.11 of a school-level standard deviation in math scale scores ($p < 0.001$), with even larger estimates of

0.14 for cumulative administrative resident exposure (Panel C) and 0.17 for Teacher Leader resident exposure (Panel D). These seem like large effects relative to the aforementioned principal effects literature, except that they are scaled in terms of school-level standard deviations, which are likely about half the size of student-level standard deviations, assuming plausible within-school student achievement correlations of about 0.2 (Bloom, n.d.). We also find positive and significant estimates for administrative resident dosage in relation to ELA scores.

However, when we turn to models in columns 7 and 8, which allow secular trends to differ within and outside of the treatment city, test score estimates are not statistically or substantively different from 0. In other words, if we believe that concurrent reforms or changes in the treatment city may have led achievement trends to differ in the treatment city for reasons unrelated to PLUS, then we would conclude that PLUS leadership did not affect test scores on average across school levels. If on the other hand, we believe that achievement trends are best captured by state averages, and that other within-city schools are not the optimal comparison group due to their sparse numbers or their unobserved attributes that led them not to be staffed by PLUS residents, then we would put greater stock in the estimates in columns 1 through 6. To err on the side of caution, we focus most of our discussion on estimates from models that include year-by-treatment city interactions. School fixed-effects estimates for the behavioral variables in columns 9 through 11 lay in the opposite of the desired direction in magnitude, though they reach statistical significance in only a few cases. Specifically, each additional administrative resident (Panel C) is associated with about a half a percentage point lower attendance rate ($p < 0.1$), and about a 4.6 percentage point higher out-of-school suspension rate ($p < 0.05$).

Relationships to suspensions are large and statistically significant for having a resident (a 10-percentage-point difference) and for cumulative number of residents (a 2.8-point difference).

The four-year graduation rate estimates are positive but pertain only to high schools. In column 12, each additional resident by year is associated with an additional 2 percentage points in four-year graduation rates ($p < 0.01$); this relationship reaches 3.6 percentage points ($p < 0.001$) for residents placed in administrative roles.

Differential Effects by School Level and Type

In Table 7, we examine whether treatment effects differ in high schools versus elementary and middle schools, and in charter versus traditional schools. For parsimony, we present results only from more conservative models in which year effects are allowed to differ between the treatment city and the rest of the state, as in columns 7 through 12 of Table 6. In Panel A of Table 7, we do see differences in PLUS effect estimates between high schools and elementary/middle schools. Specifically, an additional administrative resident-by-year has no significant association with math scores in elementary and middle schools but is linked to an additional 15 percent of a school-level standard deviation in high schools ($p < 0.001$). We also find that the relationship of administrative dosage to higher suspension rates is strongly driven by high schools, with an estimated difference of 10 percentage points ($p < 0.05$). Meanwhile, an additional Teacher Leader resident-by-year is linked to an additional 23 percent of a school-level standard deviation in high school math scores ($p < 0.001$), but to a 0.16 standard deviation lower ELA score in elementary and middle schools ($p < 0.01$). Though statistically significant, these estimates should be interpreted as descriptive, within-sample differences due to the small number of treated high schools (four) in the study.

<Insert Table 7 about here>

In Panel B, relationships of administrative leadership dosages to chronic absence rates appear desirably negative in charter schools, on the order of -4.2 percentage points ($p < 0.1$), but moderately positive in traditional schools, at about 4.9 percentage points ($p < 0.05$). The positive relationship between administrative dosages and suspension rates appears to be driven by traditional schools, with a relationship of 4 percentage points ($p < 0.01$).

Discussion

This study estimates how schools' cumulative exposure to PLUS school leadership residents relates to their subsequent school climate, achievement, behavioral, and attainment outcomes. PLUS was implemented at a time when the city was reform-focused and converting numerous schools to district-run charter schools. If we assume constant time trends across the state and treatment district, we find that PLUS resident exposure is linked to notably higher mathematics scores, chronic absences, suspensions, and graduation rates, and that this is true for residents in both administrative and Teacher Leader roles. If instead we conservatively allow time trends in achievement to vary between the state and the treatment city, we find positive relationships of administrative dosage to math achievement, suspension rates, and graduation rates in high schools. We further find that teacher leadership dosage is linked to positive math achievement in high schools and to negative ELA achievement in elementary and middle schools. But in all cases, we must interpret these estimates with caution because of the instability across models.

Because achievement is measured in school-level standard deviation units, our estimates should be divided roughly in half to be comparable to student-level standard deviation effects (Bloom, n.d.). Some larger-scale studies of alternative leadership routes (Booker & Thomas, 2014; Gates, Baird, Doss, et al., 2019) and grow-your-own models (Gates, Baird, Master, et al.,

2019) have shown larger and less-ambiguous achievement effects than those we report here, though others have also shown ambiguous and model-sensitive estimates (Corcoran et al., 2012). Moreover, effects from those studies were associated with leaders hired into principal roles rather than into subordinate roles, and they were most evident after principals had led schools for three years. In addition, all of the prior studies showed heterogeneity in effect sizes across districts. This suggests that the field would benefit from meta-analytic work examining effects of PLUS programs across cities, including the effects of leaders hired into both non-principal and principal roles.

Our estimates of the teacher leadership effects of PLUS are unstable in both direction and significance, but it is conceivable that more consistent district support for teacher leadership as part of the pipeline model could have yielded stronger effects (Authors, 2018), of the sort that Papay et al. (2016) found in their study of peer coaching. Moreover, the aforementioned literature has found long-term benefits of leadership pipeline programs in which the plan for selecting, training, hiring, evaluating, and professionally developing future principals is well-aligned across a district (Gates, Baird, Master, et al., 2019). In the current study context, the limited number of principal and administrative job openings for program residents and graduates suggests that this alignment was not fully reached. Still, future studies that examine the effects of leaders in non-principal roles would add to our understanding of how districts leverage leadership pipeline candidates before they are hired into principalships.

Though our school fixed-effects models minimize the risk of selection bias in terms of stable school characteristics, a threat to causal inference lies with any time-varying differences affecting the treatment schools differently than the comparison schools—such as changes in educator focus, effectiveness, or experience—that are not a result of PLUS residents' placement

in the schools. If the placement of administrative or teaching residents occurs due to higher-than-usual teacher turnover or to a particularly weak instructional culture, that could negatively bias the estimates, leading us to understate the effects of PLUS. On the other hand, if schools that accept residents are those that are already prioritizing instructional improvement more than other schools, that could positively bias the estimates. We know that, in a district with few job openings, residents were hired into schools that had administrator or teacher openings. These may be schools that were at least somewhat harder to staff, suggesting that the estimates, especially from the within-district models, may be too negative.

In addition, as Gates, Baird, Doss, et al. (2019) have noted, potential contamination could bias estimates conservatively toward zero. During the intervention years, the PLUS program leaders were working collaboratively with the district to improve professional development for principals across the district. The result is that most district-run schools in the city received periodic principal professional development sessions run by PLUS program staff. However, because PLUS residents received PLUS training at a much higher level of intensity than school principals who had never been part of the program, contamination effects are likely to be small.

Conclusion

Numerous studies have shown that school principals influence school effectiveness, and that the cumulative effects of principals are substantial. But there is less consensus on how to create effective principals. This is the first study we know of to attempt to estimate schoolwide effects of alternative-route school leaders working in non-principal roles. The PLUS program's logic model hypothesized that these residents could generate near-term instructional and school climate improvements even before they attained principalships. We find partial support for this theory in some models, especially for administrative roles in high schools. But we do not find

any evidence of improvement in students' attendance rates, suspension rates, or teachers' descriptions of school climate and culture. The increased suspension rate effect estimates would be consistent with a disciplinary crackdown approach to school leadership, though this was not an approach advocated by the PLUS program or the district. Instead, they promoted a restorative discipline approach that encouraged students to make amends for misbehavior rather than missing school as punishment.

Overall, the potential for time-varying alternative explanations and the small number of treatment schools make us cautious about inferring that another district that adopted a similar approach would show benefits in math performance or graduation rates, even in high schools, where our positive estimates are most robust. Rather, this study may be best understood as a description of school climate and academic trends that followed the adoption of an ambitious and intensive leadership pipeline partnership. We also offer a reminder that our estimates focus on the near-term phase of the logic model. In other studies of leadership pipeline efforts and partnerships, positive student achievement effects have emerged after leaders trained by the programs had served as principals of the same schools for at least two to three years (Booker & Thomas, 2014; Corcoran et al., 2012; Gates, Baird, Doss, et al., 2019; Gates, Baird, Master, et al., 2019). The longer-term effects of the PLUS residency program will depend on the opportunities that arise for residents to move into principalships in the city, and on what they are able to accomplish when they get there.

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.
- Bastian, K. C., & Henry, G. T. (2015). The apprentice: Pathways to the principalship and student achievement. *Educational Administration Quarterly*, 51(4), 600-639. doi: 10.1177/0013161x14562213
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust difference-in-difference estimates? *The Quarterly Journal of Economics*, 119(1), 249-275.
- Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, 41, 904-919.
- Bloom, H. (n.d.). *How should we calculate effect sizes?* New York, NY: MDRC. Retrieved from <https://www.acf.hhs.gov/sites/default/files/opre/bloom.pdf>
- Booker, K., & Thomas, J. (2014). *Impacts of New Leaders on student achievement in Oakland* (40252). Washington, DC: Mathematica Policy Research. Retrieved from <https://www.mathematica-mpr.com/our-publications-and-findings/publications/impacts-of-new-leaders-on-student-achievement>
- Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *Educational Evaluation and Policy Analysis*, 48(2), 303-333. doi: 10.3102/0002831210380788
- 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.
- Briggs, K., Cheney, G. R., Davis, J., & Moll, K. (2013). *Operating in the dark: What outdated state policies and data gaps mean for effective school leadership*. Dallas, TX: Alliance to Reform Education Leadership, George W. Bush Institute Retrieved from <https://files.eric.ed.gov/fulltext/ED560209.pdf>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50, 317-372.
- Cameron, S. V., & Heckman, J. J. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics*, 11(1, Part 1), 1-47. doi: 10.1086/298316
- Campbell, C., & Gross, B. (2012). *Principal concerns: Leadership data and strategies for states*. Seattle, WA: Center on Reinventing Public Education. Retrieved from http://www.crpe.org/sites/default/files/pub_principal_concerns_sept12.pdf
- Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value-added indicative of principal quality? *Education Finance and Policy*, 11(3), 283-309.

- Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy*, 122(2), 282-318. doi: 10.1086/675238
- Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 31(1), 92-109. doi: 10.1016/j.econedurev.2011.09.001
- Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111(1), 180-213.
- Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2009). *The New York City Aspiring Principals Program: A school-level evaluation*. New York: Institute for Education and Social Policy, New York University.
- Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012). Training your own: The impact of New York City's aspiring principals program on student achievement. *Educational Evaluation and Policy Analysis*, 34(2), 232-253.
- Darling-Hammond, L., LaPointe, M., Meyerson, D., Orr, M. T., & Cohen, C. (2007). *Preparing school leaders for a changing world: Lessons from exemplary leadership development programs*. Stanford, CA: Stanford University, Stanford Educational Leadership Institute. Retrieved from <http://seli.stanford.edu>
- Davis, B. W., Gooden, M. A., & Bowers, A. J. (2017). Pathways to the principalship: An event history analysis of the careers of teachers with principal certification. *American Educational Research Journal*, 54(2), 207-240. doi: <https://doi.org/10.3102/0002831216687530>
- Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics*, 47(2), 634-663.
- Fryer, R. G., Jr. (2017). *Management and student achievement: Evidence from a randomized field experiment* (23437). Cambridge, MA: National Bureau of Economic Research. Retrieved from www.nber.org/papers/w23437
- Gates, S. M., Baird, M. D., Doss, C. J., Hamilton, L. S., Opper, I. M., Master, B. K., . . . Zaber, M. A. (2019). *Preparing school leaders for success: Evaluation of New Leaders' Aspiring Principals Program, 2012-2017*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR2812.html
- Gates, S. M., Baird, M. D., Master, B. K., & Chavez-Herrerias, E. R. (2019). *Principal pipelines: A feasible, affordable, and effective way for districts to improve schools*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR2666.html
- Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., . . . Gu, K. (2014). *Preparing principals to raise student achievement: Implementation and effects of the New Leaders program in ten districts*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR507.html

- Goddard, R. D., Salloum, S. J., & Berebitsky, D. (2009). Trust as a mediator of the relationships between poverty, racial composition, and academic achievement: Evidence from Michigan's public elementary schools. *Educational Administration Quarterly*, 45(2), 292-311. doi: 10.1177/0013161x08330503
- Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*, 109(4), 877-896.
- 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., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 37(1), 3-28. doi: 10.3102/0162373714523831
- Grissom, J. A., Mitani, H., & Woo, D. S. (2018). Principal preparation programs and principal outcomes. *Educational Administration Quarterly*. doi: 10.1177/0013161x18785865
- Hassel, E. A., Hassel, B. C., Arkin, M. D., Kowal, J. M., & Steiner, L. M. (2006). *School restructuring under No Child Left Behind: What works when? A guide for school leaders*. Washington, DC: The Center for Comprehensive School Reform and Improvement, Learning Point Associates. Retrieved from <https://files.eric.ed.gov/fulltext/ED496104.pdf>
- Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *IES Practice Guide: Turning Around Chronically Low-Performing Schools* (NCEE 2008-4020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Herrington, C. D., & Wills, B. K. (2005). Decertifying the principalship: The politics of administrator preparation in Florida. *Educational Policy*, 19(1), 181-200. doi: 10.1177/0895904804271715
- Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85-108. doi: 10.1257/app.1.4.85
- Jacob, R., Goddard, R. D., Kim, M., Miller, R., & Goddard, Y. L. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, 37(3), 314-332.
- Kane, T. J., & Staiger, D. O. (2005). *Using imperfect information to identify effective teachers*. Cambridge, MA: National Bureau of Economic Research.

- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 0*(0). doi: 10.3102/0034654318759268
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. St. Paul, MN: Center for Applied Research and Educational Improvement, University of Minnesota. Retrieved from <https://conservancy.umn.edu/bitstream/handle/11299/2035/CAREI?sequence=1>
- Levine, A. (2005). *Educating school leaders*. Washington, DC: The Education Schools Project.
- National Alliance for Public Charter Schools. (2008). *Charter school executives: Toward a new generation of leadership*. Washington, DC: National Alliance for Public Charter Schools. Retrieved from https://www.publiccharters.org/sites/default/files/documents/2017-11/2008_Toward_a_New_Generation_of_Leadership_20110402T222340.pdf
- National School Climate Council. (2007). *The school climate challenge: Narrowing the gap between school climate research and school climate policy, practice guidelines and teacher education policy*. New York: National School Climate Council. Retrieved from <http://community-matters.org/downloads/school-climate-challenge-policy-paper.pdf>
- Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. (2016). *Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data*. National Bureau of Economic Research.
- Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics, 1*-14. doi: 10.1080/09645292.2018.1484426
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal, 50*(1), 4-36. doi: 10.3102/0002831212463813
- Steele, J. L., Steiner, E. D., & Hamilton, L. S. (2018). *Distributing instructional leadership: Implementation lessons from an urban school leadership residency program (WR-1275-NTP)*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/working_papers/WR1275.html

Table 1. Initial placements and first-year completers, by cohort and initial role

Cohort	Start Year	Apprentice School							
		Lead Educator		Leader (Charter)		Teacher Leader		Total	
		I	C	I	C	I	C	I	C
1	2014-15	8	6	1	1	0	0	9	7
2	2015-16	2	2	4	3	7	6*	13	11
3	2016-17	0	0	2	2	6	5*	8	7
Total		10	8	7	6	13	11	30	25

I=Initial Placement, C=First-Year Completer

* One each in Cohorts 2 and 3 were charter school placements.

Table 2. Teacher survey composites, source items, means, and standard deviations

Composite	Mean	SD	# of Items	Cronbach's α	Description
School Climate	7.2	1.9	34	0.95	Teacher-by-time-period mean of the other 4 composites
Learning Environment and Leadership	7.0	2.3	8	0.94	Learning environment is safe, with consistent expectations. The school leader sets and executes priorities for the school, seeks feedback, and works to retain effective teachers.
Peer Collaboration for Student Growth	7.4	1.8	9	0.90	Expectations for quality teaching are shared among teachers and reinforced through collaboration, who also collaborate on use of data to make instructional decisions.
Observation, Feedback and Professional Development	7.2	2.1	10	0.93	Opportunities to improve instructional practice are accessible and high-quality, including frequent and useful feedback from observations.
Fair Evaluation	6.9	2.2	7	0.92	Evaluations of teaching are accurate, useful, and reflect clear expectations.
Years Plan to Stay	6.4	3.8	1	NA	Years the teacher plans to stay in the current school, from 0 to 10+, rescaled as 0 to 10.

Table 3. Descriptive statistics for ever-treated and never-treated schools at program inception in 2014-15

	n Ever-treated	Mean Ever-treated	n Never-treated Treatment City	Mean Never-treated Treatment City	Mean Difference Treatment City	p-value Treatment City	n Never-treated Statewide	Mean Never-treated Statewide	Mean Difference Statewide	p-value Statewide
<i>Control Variables</i>										
Element./K-8	13	0.538	24	0.792	-0.253	0.11	2465	0.626	-0.088	0.52
Middle	13	0.154	24	0.042	0.112	0.25	2465	0.181	-0.027	0.80
High	13	0.231	24	0.083	0.147	0.22	2465	0.166	0.064	0.53
Other	13	0.077	24	0.083	-0.006	0.95	2465	0.027	0.050	0.27
Charter	13	0.385	24	0.458	-0.074	0.68	2465	0.035	0.349	0.00
Enrollment % Free/Red. Lunch	13	490.0	24	410.0	79.2	0.33	2465	552.0	-62.3	0.58
% Asian	13	93.3	24	89.8	3.5	0.20	2462	36.1	57.1	0.00
% Black	13	0.6	24	0.8	-0.3	0.62	2465	8.9	-8.4	0.02
% Hispanic	13	45.8	24	49.0	-3.1	0.66	2465	16.3	29.5	0.00
% White	13	53.1	24	49.5	3.7	0.61	2465	23.9	29.3	0.00
Stu/Tch Ratio	13	0.5	24	0.7	-0.2	0.39	2465	50.9	-50.4	0.00
<i>Outcome Variables</i>										
Math scores	13	10.9	24	10.6	0.3	0.72	2462	12.7	-1.8	0.57
Math scores	13	-1.790	21	-1.490	-0.296	0.29	2204	0.023	-1.810	0.00
ELA scores	13	-2.160	21	-1.500	-0.661	0.03	2200	0.082	-2.240	0.00
Math SGP	10	30.9	18	37.3	-6.4	0.37	1727	50.7	-19.8	0.00
ELA SGP	10	32.6	18	44.3	-11.7	0.03	1733	51.1	-18.4	0.00
Math % prof+	13	3.5	21	12.2	-8.8	0.04	2204	43.6	-40.1	0.00
ELA % prof+	13	5.8	21	17.7	-12.0	0.02	2200	58.0	-52.2	0.00
% Chron. Abs. Attendance Rate	6	19.8	14	18.6	1.2	0.84	1791	9.1	10.7	0.01
% Suspended	12	90.3	14	91.6	-1.3	0.44	14	91.6	-1.3	0.44
4-yr Grad Rate	13	20.6	20	14.4	6.1	0.30	2392	4.5	16.0	0.00
4-yr Grad Rate	2	55.0	4	96.8	-41.8	0.00	397	90.6	-35.6	0.00

Table 4. Estimated effects of resident placement/dosage on composite teacher reports of school climate

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	School Climate	Learning Environ & Leadership	Peer Collab. for Stu. Growth	Observati on, Feedback, & PD	Fair Evaluatio n	Years Plan to Stay
PANEL A						
Had 1+ Residents	0.0333 (0.214)	0.0671 (0.296)	0.0378 (0.190)	0.0386 (0.188)	-0.133 (0.171)	-0.0724 (0.153)
PANEL B						
Cumulative number of residents over time	-0.0365 (0.0636)	-0.0558 (0.0950)	-0.0372 (0.0453)	-0.0330 (0.0667)	-0.0219 (0.0642)	0.00351 (0.0750)
PANEL C						
Cumulative number of administrative residents	-0.0183 (0.0768)	-0.0247 (0.108)	-0.0397 (0.0587)	0.00820 (0.0900)	0.0137 (0.0809)	-0.0963 (0.0924)
PANEL D						
Cumulative number of Teacher Leader residents	-0.0985 (0.128)	-0.156 (0.208)	-0.0616 (0.0827)	-0.135 (0.120)	-0.106 (0.138)	0.192 (0.115)
Observations	6,947	6,337	6,401	6,282	6,566	6,823
Number of schools	30	30	30	30	30	30
Avg. obs. w/in schools	231.6	211.2	213.4	209.4	218.9	227.4
Intraclass Corr: Has Residents	0.290	0.389	0.277	0.267	0.213	0.0679
R-sq_Has Residents	0.064	0.059	0.056	0.064	0.0534	0.007
School Fixed Effects	x	x	x	x	x	x

Cluster-robust standard errors in parentheses. Models include controls for schools' total enrollment size, percent free/reduced-price meal eligible, racial/ethnic student percentages, and student-teacher ratios. R-squared in fixed-effect models is within-schools.

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10

Table 5. Estimated differential effects of treatment variables on school climate measures for high schools (Panel A) and charter school (Panel B)

PANEL A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	School Climate	Learning Environ & Leadership	Peer Collab. for Stu. Growth	Observation, Feedback, & PD	Fair Evaluation	Years Plan to Stay	School Climate	Learning Environ & Leadership	Peer Collab. for Stu. Growth	Observation, Feedback, & PD	Fair Evaluation	Years Plan to Stay
High school	0.0625 (0.451)	0.392 (0.623)	-0.0105 (0.315)	0.0684 (0.412)	-0.156 (0.441)	0.386 (0.376)	0.0581 (0.465)	0.420 (0.652)	-0.0217 (0.325)	0.0686 (0.418)	-0.205 (0.447)	0.405 (0.375)
Cumulative administrative residents	-0.0560 (0.104)	-0.0949 (0.143)	-0.0574 (0.0802)	-0.0313 (0.112)	-0.0139 (0.113)	-0.0700 (0.111)						
Cumulative administrative residents * high school	-0.0202 (0.115)	0.0288 (0.183)	-0.0498 (0.0916)	-0.0342 (0.149)	-0.0597 (0.118)	-0.0581 (0.133)						
Cumulative Teacher Leader residents							-0.0483 (0.154)	-0.0727 (0.243)	-0.0133 (0.0990)	-0.0800 (0.144)	-0.0942 (0.176)	0.248* (0.101)
Cumulative Teacher Leader residents * high school							-0.370* (0.163)	-0.489~ (0.275)	-0.346** (0.122)	-0.393** (0.137)	-0.186 (0.183)	-0.256 (0.188)
Observations	6,947	6,337	6,401	6,282	6,566	6,823	6,947	6,337	6,401	6,282	6,566	6,823
Number of schools	30	30	30	30	30	30	30	30	30	30	30	30
Intraclass correlation	0.0663	0.0978	0.0528	0.0475	0.0439	0.0144	0.0854	0.132	0.0660	0.0538	0.0592	0.0149
PANEL B	(1)	(2)	(3)	(4)	(5)	(6)						
Charter	-0.204 (0.208)	-0.712~ (0.396)	-0.180 (0.146)	0.00762 (0.249)	0.293 (0.286)	-1.385*** (0.386)						
Cumulative admin. residents	0.00141 (0.101)	0.0164 (0.147)	-0.0265 (0.0725)	0.00829 (0.117)	3 (0.0911)	-0.0473 (0.0901)						
Cumulative admin. residents * Charter	-0.217* (0.0966)	-0.319* (0.153)	-0.166* (0.0826)	-0.175 (0.123)	-0.125 (0.115)	-0.154 (0.124)						
Observations	6,947	6,337	6,401	6,282	6,566	6,823						
Number of schools	30	30	30	30	30	30						
Intraclass correlation	0.0727	0.115	0.0593	0.0500	0.0465	0.0145						

Cluster-robust standard errors in parentheses. Random-effect models include controls for school charter status, total enrollment size, percent free/reduced-price meal eligible, racial/ethnic student percentages, student-teacher ratios, a treatment-city indicator, and an indicator of whether the school ever received a resident.

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10

Table 6. Estimated effects of residency placement/dosage on school-level achievement, behavior, and attainment measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Math Scores	ELA Scores	% Chron abs	Attend. rate	Suspen sion rate	4-year grad rate	Math Scores	ELA Scores	% Chron abs	Attend. rate	Suspen sion rate	4-year grad rate
PANEL A												
Had 1+ Residents	0.356*** (0.0654)	0.173 (0.144)	2.345 (1.633)	-0.241 (0.773)	9.450~ (4.874)	3.666 (5.937)	0.0288 (0.109)	-0.0087 (0.151)	2.009 (1.510)	-0.241 (0.773)	10.84* (4.988)	5.765 (4.953)
PANEL B												
Cumulative number of residents over time	0.108*** (0.0221)	0.0509 (0.0387)	0.543 (0.413)	-0.226 (0.267)	2.227~ (1.176)	1.209 (0.945)	0.0009 (0.0353)	-0.0079 (0.0450)	0.0635 (0.464)	-0.226 (0.267)	2.858* (1.183)	2.016** (0.742)
PANEL C												
Cumulative number of administrative residents	0.141*** (0.0343)	0.0980* (0.0467)	1.136~ (0.625)	-0.560~ (0.312)	3.819~ (2.104)	2.527* (1.083)	0.0050 (0.0432)	0.0346 (0.0542)	0.746 (0.631)	-0.560~ (0.312)	4.594* (2.204)	3.627*** (0.853)
PANEL D												
Cumulative number of Teacher Leaders	0.172*** (0.0495)	-0.0007 (0.0753)	0.182 (0.840)	0.092 (0.500)	1.237 (1.127)	1.019 (2.825)	-0.0084 (0.0522)	-0.118~ (0.0701)	-0.943 (0.902)	0.092 (0.500)	1.042 (1.608)	2.453 (2.347)
Observations	26,090	26,087	9,812	106	29,676	2,367	26,090	26,087	9,812	106	29,676	2,367
Number of schools	2,399	2,397	2,265	30	2,679	421	2,399	2,397	2,265	30	2,679	421
Avg. obs. w/in schools	10.88	10.88	4.332	3.533	11.08	5.622	10.88	10.88	4.332	3.533	11.08	5.622
Intraclass Corr: Has Residents	0.764	0.796	0.675	0.881	0.652	0.935	0.761	0.794	0.676	0.881	0.646	0.935
R-sq_Has Residents	0.054	0.044	0.015	0.214	0.024	0.120	0.058	0.046	0.015	0.214	0.026	0.123
School Fixed Effects	x	x	x	x	x	x	x	x	x	x	x	x
Year-by-Treatment District Interactions							x	x	x	x	x	x

Cluster-robust standard errors in parentheses. Models include controls for schools' total enrollment size, percent free/reduced-price meal eligible, racial/ethnic student percentages, and student-teacher ratios. R-squared in fixed-effect models is within-schools. Columns 4 and 10 are identical because attendance rates are available only in the treatment district.

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10

Table 7. Estimated differential effects of residency placements for high schools (Panel A) and for charter schools (Panel B)

PANEL A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Math Scores	ELA Scores	% Chron abs	Attend. rate	Suspension rate	Math Scores	ELA Scores	% Chron abs	Attend. rate	Suspension rate
High school	-0.0924*	-0.0380	6.331***	-1.681	9.426***	-0.0924*	-0.0377	6.329***	-1.997	9.450***
	(0.0442)	(0.0543)	(0.721)	(1.886)	(0.710)	(0.0443)	(0.0544)	(0.725)	(1.859)	(0.710)
Cumulative administrative residents	-0.0604	-0.0444	1.099	-0.378~	0.503					
	(0.0491)	(0.0606)	(0.685)	(0.219)	(1.723)					
Cumulative administrative residents * high school	0.149***	0.115	3.336	-0.789	10.16*					
	(0.0403)	(0.0931)	(4.136)	(0.710)	(4.299)					
Cumulative Teacher Leader residents						-0.0378	-0.159**	-0.691	-0.0018	1.523
						(0.0401)	(0.0530)	(0.744)	(0.218)	(1.292)
Cumulative Teacher Leader residents * high school						0.227***	0.279	6.285	-0.0599	0.471
						(0.0380)	(0.173)	(5.352)	(1.868)	(3.651)
Observations	26,090	26,087	9,812	106	29,676	26,090	26,087	9,812	106	29,676
Number of schools	2,399	2,397	2,265	30	2,679	2,399	2,397	2,265	30	2,679
Intraclass correlation	0.644	0.689	0.464	0.835	0.501	0.644	0.689	0.461	0.616	0.500
PANEL B	(1)	(2)	(3)	(4)	(5)					
	Math Scores	ELA Scores	% Chron abs	Attend. rate	Suspension rate					
Charter	0.292***	0.337***	-3.266**	0.223	-0.862					
	(0.0650)	(0.0680)	(1.249)	(0.483)	(0.981)					
Cumulative administrative residents	0.0217	-0.0219	4.850*	-0.797	4.094**					
	(0.0591)	(0.0926)	(2.373)	(0.506)	(1.541)					
Cumulative administrative residents * charter	-0.0767	0.00783	-4.240~	0.335	-0.557					
	(0.0615)	(0.102)	(2.350)	(0.414)	(4.046)					
Observations	26,090	26,087	9,812	106	29,676					
Number of schools	2,399	2,397	2,265	30	2,679					
Intraclass correlation	0.644	0.689	0.463	0.816	0.500					

Cluster-robust standard errors in parentheses. Random-effect models include controls for total enrollment size, percent free/reduced-price meal eligible, racial/ethnic student percentages, and student-teacher ratios. All models shown include year-by-treatment district interactions, corresponding to Table 6 columns 7 through 12.

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.10

Figure 1. Logic model showing anticipated near-term and mid-term results of school leadership residency program

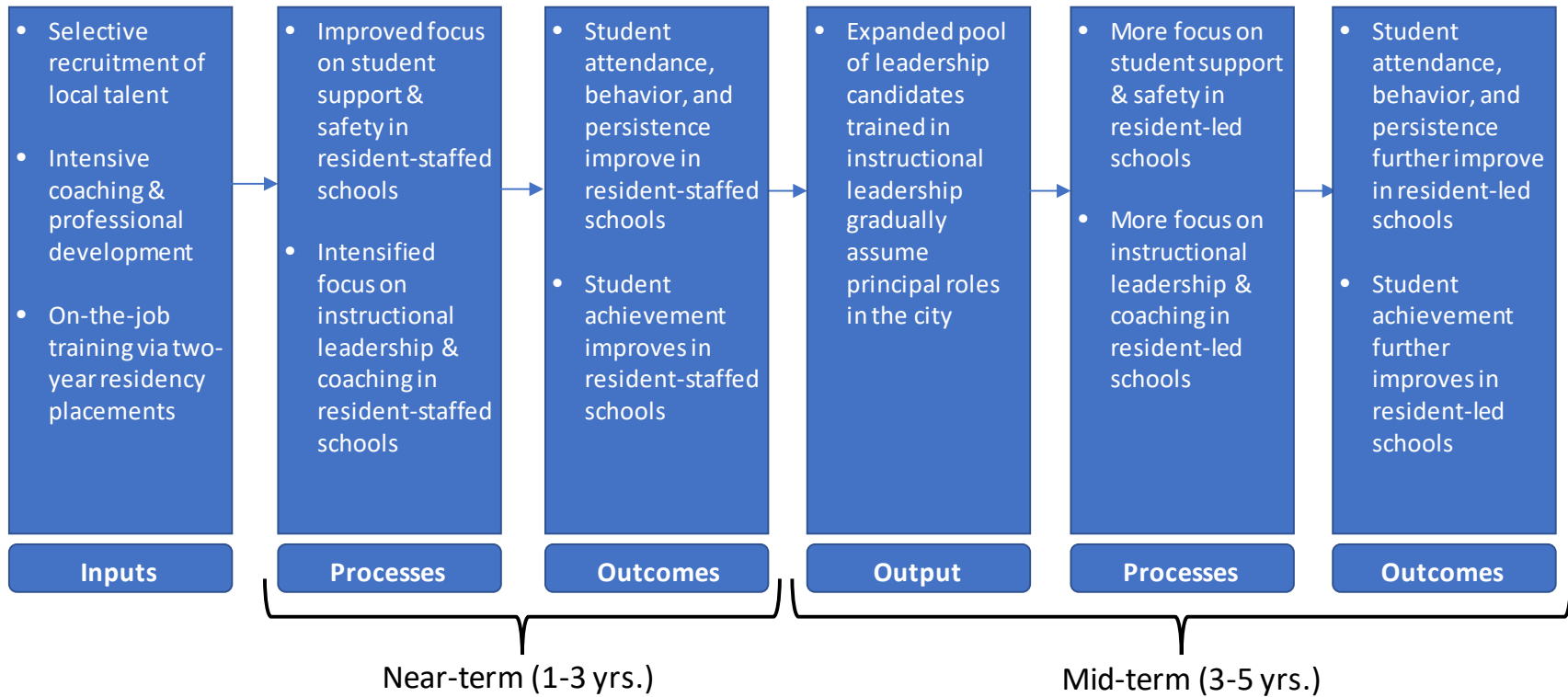


Figure 2. Time trends in school climate composite variables (range: 0-10) for ever-treated (n=13) versus never-treated (n=24) schools in the treatment city

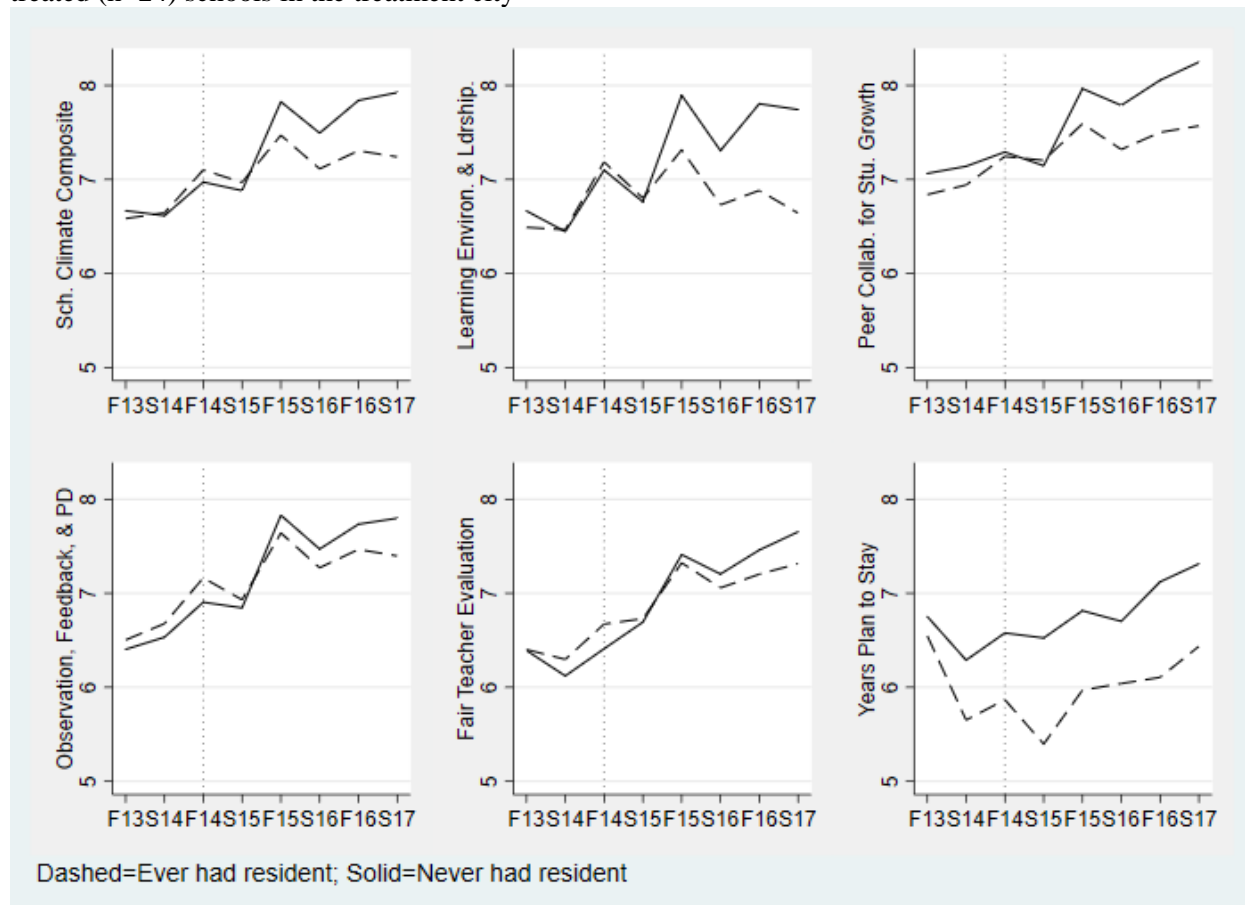
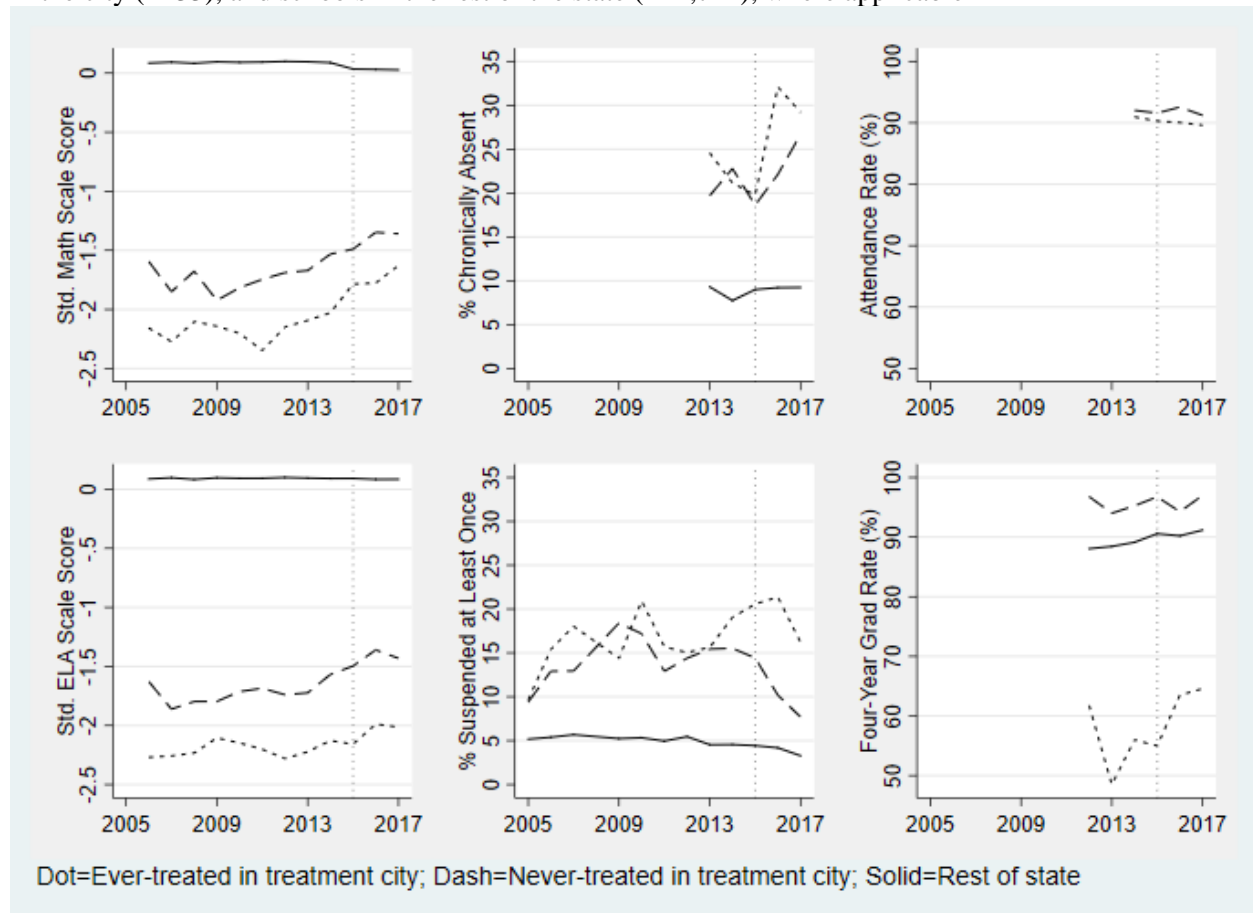


Figure 3. Time trends in academic and behavioral outcomes in ever-treated schools (n=13), never-treated schools in the city (n=33), and schools in the rest of the state (n=2,722), where applicable



Note: Y-axis scales differ among graphs.