WORKING
PAPER

Preliminary Findings from the New Leaders for New Schools Evaluation

FRANCISCO MARTORELL, PAUL HEATON,
SUSAN M. GATES, LAURA S. HAMILTON

WR-739-NLNS

February 2010

EDUCATION

ABSTRACT

Effective school leadership is widely seen as a key determinant of student achievement, yet it remains unclear what constitutes an effective principal. To address the need to develop new principals to lead urban schools, the New Leaders for New Schools organization was established with the goal of ensuring high academic achievement for all students by attracting, preparing, and supporting leaders for urban public schools. This working paper presents preliminary findings on the impact of attending a school led by a K-8 school led by a New Leader. Using longitudinal student-level data collected from the six cities in which New Leaders had placed principals by the 2007-08 school year, we attempt to estimate the effect of attending a school led by a New Leader using panel data methods to mitigate biases from nonrandom sorting of students and principals to schools. The estimates suggest that there is a positive association between achievement and having a New Leader in his or her second (or higher) year of tenure, while there is a small negative relationship between achievement and attending a school led by a first-year New Leader.

PREFACE

New Leaders for New Schools ("New Leaders") is dedicated to promoting student achievement by developing outstanding school leaders to serve in urban schools. RAND is conducting a formative and a summative evaluation of the NLNS program, its theory of action, and its implementation.

This paper presents interim results from the third year of analysis of student achievement data. The analysis compares schools with a New Leaders principal to those without a New Leaders principal. A final report will be available at the end of the longitudinal study, anticipated by 2014.

The extended research period for the evaluation is driven by three major factors. First, New Leaders' unit of impact is a school, and schools are dispersed throughout the country. As a result, the numbers of schools for any one level, any one city, and any one type (charter, start-up) remain even now quite small. Effectively assessing program impact is best accomplished when the numbers are large enough to detect systematic effects. Second, the New Leaders program underwent substantive changes beginning in the 2007-2008 school year, both requiring a significantly longer commitment from the entering aspiring principals and focusing the work on reaching specific student achievement targets. Any impact from these important programmatic shifts we anticipate will be evident across a large enough pool of schools in 2010 or 2011, depending on the numbers of principals and their tenure in those years. The final reason is purely logistical: student-level achievement data are not available across all districts until as much as a year after the tests are taken, and analyzing the data requires time.

This research has been conducted in RAND Education, a unit of the RAND Corporation, under a contract with New Leaders for New Schools.

CONTENTS

TABLES

GLOSSARY

Symbol	Definition
BCPS	Baltimore City Public Schools
CPS	Chicago Public Schools
DCPS	Washington DC Public Schools
MCS	Memphis City Schools
NYC	New York City Schools
OUSD	Oakland Unified School District
PSAE	Prairie State Achievement Exam

1. INTRODUCTION

The view that effective principals are crucial for a school's success is widespread (Brewer, 1993; Teske and Schneider, 1999). The recent emphasis on school-level accountability implies that school-level decisions matter a great deal. Many of these decisions are made or substantially influenced by the school principal. For example, principals exert considerable influence over the composition of the teacher and support workforce, school culture, instruction as well as professional development, and supports to promote teacher effectiveness. Despite conventional wisdom that principals and the decisions they make are critical to school outcomes, there is little consensus as to the specific attributes that characterize an effective principal or even how much principals actually affect student outcomes. Although several recent studies have examined whether factors such as principal experience "matter" (Clark, Martorell, and Rockoff, 2009; Branch, Hanushek, and Rivkin, 2009; Coelli and Green, 2009), and some research has identified principal practices that are associated with student achievement gains (Leithwood et al., 2004; Waters et al., 2003), it is still not clear whether it is possible to systematically identify who will be a good principal from a given pool of applicants. Moreover, there is little evidence regarding whether and how training and support for new principals enhances their effectiveness.

One organization which has sought to improve principal effectiveness through selection, training, and support is New Leaders for New Schools ("New Leaders"). The goal of the New Leaders program is to ensure high academic achievement for all students by attracting, preparing, and supporting leaders for urban public schools. Core elements of the New Leaders program are a rigorous outreach and selection process, a summer learning experience (Foundations), a year-long structured residency experience, placement assistance, and post-residency training and support. New Leaders' theory of action posits that effective school leaders will drive effective teaching which will in turn lead to significant gains in student achievement (New Leaders for New Schools, 2009). Effective school leaders focus on five categories of actions.

Within each category, the specific actions and emphasis must be tailored to the school context or "stage". The ability to diagnose the stage of a school is itself a critical school leadership competency. The five categories of actions emphasized by New Leaders are: ensuring effective, standards-based teaching and learning (includes data driven decision-making), promoting a culture focused on student achievement, aligning staff to school goals, ensuring effective operations and personal leadership (New Leaders for New Schools, 2009). New Leaders initiated its training program in the fall of 2001, and since then nearly 600 New Leaders had been placed in various school and district leadership positions across the partner districts.

This paper presents preliminary findings evaluating the impact New Leaders principals have on student achievement. Using longitudinal student-level data collected from the six cities in which New Leaders had placed principals by the 2007-08 school year, we attempt to estimate the effect of attending a school led by a New Leader on standardized test scores. New Leaders are assigned to schools in a highly non-random fashion, and often lead schools that serve some of the more disadvantaged segments of the school districts in which they are placed. This poses considerable methodological hurdles for program evaluation, as it is difficult to isolate the effect of the New Leaders from the influence of confounding factors. In this paper, we outline the nature of this challenge in greater detail as well as describe the statistical approaches we use to confront it. Specifically, we take advantage of the longitudinal information on students to estimate, when possible, panel-data models that control for permanent unobserved heterogeneity.

We begin with a brief overview of the New Leaders program in Section 2. Next, we describe the data we have collected for this project in Section 3. Section 4 discusses the econometric strategy. Section 5 describes the results we have found. Finally, Section 6 concludes the paper.

2. DESCRIPTION OF THE NEW LEADERS PROGRAM

The New Leaders organization was established in 2000 and began training principals in the fall of 2001. Thirteen individuals completed the residency program and were placed in school leadership positions in three cities or regions the following fall: New York, Chicago, and the San Francisco Bay Area. Since that time, New Leaders has expanded the number of residents as well as its regional coverage through a formal district selection process designed to identify partners who are committed to improving urban school leadership. As of 2010, New Leaders had 90 residents in twelve urban centers: New York city, Newark, NJ, Chicago, California's Bay Area, Washington DC, Memphis, Baltimore, Prince George's County, Milwaukee, Jefferson Parish, New Orleans, and Charlotte.

As mentioned above, the New Leaders program includes outreach and selection, summer Foundations, a year-long structured residency experience (described below), placement assistance, and post-residency training and support once they are placed as principals. New Leaders engages in substantial outreach to identify and encourage potential applicants. The selection process is highly competitive and guided by a set of selection criteria that is informed by current leadership research. These criteria include, first and foremost, a belief that all children can excel academically. Other selection criteria include personal responsibility and self-motivation, knowledge of teaching and learning, results orientation, strategic management, and leadership characteristics.[1] The final admissions hurdle is a finalist selection day where applicants engage in and are judged on a range of activities similar to those that a principal undertakes. Individuals who are accepted into the program pay no tuition and receive a salary and benefits from the partner district or charter school organization during their full-time training experience. In exchange, they must commit to serving as a school leader for six years, including the residency year.

[1] Current selection criteria can be found on the New Leaders website at http://www.nlns.org/Criteria.jsp. The criteria have been refined over time.

Aspiring principals from all cities attend a summer Foundation, which provides intensive coursework and group activities designed to prepare individuals for their residency experience. The residency occurs in a school in New Leaders partner districts. The resident is mentored by both a Leadership Coach (employed by New Leaders) and by the principal of the building in which the resident is placed (also known as the "mentor principal").[2] Over the course of the year, the resident engages in three year-long projects designed to develop the resident's proficiency in the New Leaders leadership competencies. Progress in the development of these competencies is assessed throughout the year. Based on Resident progress, New Leaders provides a recommendation to the district or charter partner about the readiness of the Resident to take on the principalship. All New Leaders Residents must go through the existing selection and hiring processes for the partner district or charter school organization. The typical placement is a principalship, but may also be an assistant principalship either because the Resident is deemed not yet ready for the principalship or because of a lack of principal vacancies, or a "planning year" for individuals who will start up a new school. After placement, New Leaders receive ongoing support from Leadership Coaches.

New Leaders continuously looks for opportunities to improve its selection, training, and support activities in response to formative feedback and outcomes analysis. While many of these changes have been minor or incremental, in 2007 New Leaders engaged in a major revision to its mission, vision, and strategy. The original stated mission of New Leaders was to *promote* high achievement for all children, and the goal was to recruit a specified number of aspiring principals. New Leaders has since changed its mission statement from promoting to *ensuring* high achievement. Goals are stated in terms of student outcomes (By 2014, 60% of schools led by New Leaders principals will be on track to having 90 to 100-percent of students achieve proficiency in core academic subjects by the principal's fifth year. By 2014, 90% of high schools will be on

[2] It is possible that the presence of a resident principal could affect student achievement by easing the burden on the regular principal. In future work, we plan on examining this possibility.

track to having 90 to 100-percent of students graduate from high school). These revisions to mission, vision and strategy coincided with the development and subsequent dissemination of the Urban Excellence Framework™ (UEF)™[3]. The UEF™ identifies leadership actions that can drive effective teaching and lead to breakthrough gains in student achievement in urban schools and provides a conceptual framework for understanding how leadership actions matter in different school settings. It is a conceptual framework that provides clarity and focus to the theory of action that New Leaders has followed since its inception. New Leaders is using the UEF™ to shape refinements to specific program elements, especially the foundations curriculum and residency experience. These modifications do not reflect a drastic change in the nature of the treatment as much as a continuous trajectory of refinement to that treatment -- with a potentially large refinement experienced by the cohort of principals who were residents in the 2008-2009 school year. New Leaders organization is disseminating the UEF™ insights in the hopes of influencing decisions made by the federal government, state governments, school districts and private funders in support of these leadership actions in all urban schools (New Leaders for New Schools, 2009).

[3] http://www.nlns.org/documents/uef/principal_effectiveness_nlns.pdf

3. DATA

STUDY SITES

This analysis focuses on New Leaders placed in the 2007-08 school year or earlier. By this time, New Leaders had been placed in six regions: New York City, Chicago, the San Francisco Bay Area, Washington DC, Memphis, and Baltimore. We limited our analysis to the school district in each region where most New Leaders principals had been placed. Specifically, the districts from which we requested data are Memphis City Schools (MCS), Chicago Public Schools (CPS), New York City (NYC) Public Schools, Washington DC Public Schools (DCPS), Baltimore City Public Schools (BCPS), and the Oakland Unified School District (OUSD).

We requested student- and school-level data from each of these districts. In addition, in some districts we had to make separate requests for data on charter schools. In Washington DC, we requested data from charter schools administered by the DC Public Charter School Board and in Oakland, we requested data from the two charter schools that had New Leaders principals.[4]

STUDENT-LEVEL DATA

Our study uses standardized test scores to measure student achievement. We focus on math and reading scores because every district administers tests in these subjects, and does so in all grades in which they administer tests in any subject (in contrast, tests in subjects such as social studies are commonly offered in fewer grades). Moreover, these are the primary subjects for which all public school principals are being held accountable under No Child Left Behind. In four districts (MCS, OUSD, BCPS, and DCPS), the test that we use to analyze achievement in K-8 is also the statewide test used for NCLB purposes. In Chicago, we use a district-administered test (the Iowa Test of Basic Skills,

[4] The most important omission are schools in the Bay Area that are outside of OUSD. These are excluded because they are located in disparate districts throughout the region and we do not have access to data on students in comparison schools. An ongoing data collection effort is to get student-level data.

henceforth ITBS) prior to the 2005-06 school year, because this test was administered in grades 3 through 8, while the statewide test was only given in grades 3, 5, and 8.[5] In New York, prior to the 2005-06 school year, the student achievement data come from a state-administered test given in grades 4 and 8 and a city-administered test in grades 3, 5, 6, and 7. Because the tests vary over time in a given city, we produced standardized scores that are comparable across years and grades. Specifically, percentile ranks were computed for the unadjusted scale scores by grade and school year (separately for each district) and converted into percentile ranks using normal curve equivalents.[6] Strictly speaking, this standardization when the tests used change in a district is valid under the strong assumption that tests are linearly equitable. However, even if this is not true, it is the only way that longitudinal analyses can be conducted. In addition, it is not clear that the distortion introduced in the standardization would be differential across New Leaders or non-New Leaders schools.

New Leaders have also been placed in high schools where testing often occurs in only one grade (in most cities). This paper focuses on elementary and middle school students. The analysis at the high school level faces several important methodological challenges, which we are addressing in ongoing work.

As we explain below, our analytic strategy relies on within-student variation in exposure to a New Leaders-led school or controlling for achievement prior to entering a New Leaders-led school. Thus, it is important to obtain access to data prior to the placement of New Leader in a district. For all cities, we have obtained at least two years of "pre-program" data. Furthermore, these data files have individual student pseudo-identifiers which are used to generate longitudinal student records.[7]

---

[5] Starting in the 2005-06 school year, CPS only administered the statewide test, the Illinois Standards Assessment Test (ISAT). However, starting in that year, the statewide test was administered in grades 3 through 8.

[6] These standardizations and the estimation of treatment effects are done separately by city (see below for a description of the statistical models).

[7] An important limitation of the data in OUSD is that the student identifiers for charter schools are not the same as the identifiers for

In addition to the achievement data, we also obtained student-level demographic characteristics (e.g., race) and information on participation in school services such as free/reduced lunch. Appendix Table A.1 lists the variables obtained for each city. Such information is useful as controls for differences between students in New Leaders and non-New Leaders schools.

INFORMATION ON PRINCIPALS

The New Leaders program places principals with little or no previous experience as a school principal. Because recent research has found that principals become more effective as they acquire experience (Clark, Martorell, and Rockoff, 2009; Branch, Hanushek, and Rivkin, 2009), a concern is that the effect of the New Leaders program could be biased downward. To address this issue, we collected information from each district on principal experience.

The quality and coverage of the principal experience data varies considerably across cities (see Appendix B for details). In some cases (e.g., New York City regular district schools), the data are fairly complete, and when compared to self-reported data collected for a survey, appear to be quite reliable. In other cases, such as Baltimore, the data are missing for a relatively large fraction of the schools. Appendix C discusses a sensitivity analysis where we estimated the models for MCS, a district where the principal data are quite detailed, with and without principal covariates.

SCHOOL-LEVEL DATA AND INFORMATION ON NEW LEADERS PROGRAM STATUS

The school-level data provided by the school districts generally consisted of the range of grades at the school as well as the school's type (i.e., charter or regular). Because there is evidence that the demographic makeup of a school affects student achievement above and beyond the influence of a student's own characteristics (Card and Rothstein, 2006), we also control for a school's demographic

---

regular district schools. This makes it impossible to track students as they leave a regular OUSD school to enroll in a charter school (and vice-versa). Below we describe how this is handled in the estimation.

composition. We received this type of information from CPS. In the other sites, we created school-level variables directly from the student data.

New Leaders provided us with information on each principal trained by the organization who was placed in a school in one of the six study sites. These data included the starting and ending date of the principal's tenure with the school as well as the school name. We matched this information to the student- and school-level datasets, and created variables indicating the number of years a school had had a New Leaders principal.

COUNTS OF SCHOOLS IN THE ANALYSIS

Table 3.1 reports the number of program schools by district that are included in the student achievement analysis. These counts are done by the years of tenure a principal has at a school.[8] Schools were only included in a district's count if they contributed to the student achievement program effect estimates.[9] On the other hand, these counts include schools that are no longer led by New Leaders principals. Note that schools that contribute to the "At least two years" row also contribute to the rows for "At least one year" row, and similarly for other years, so that the total number of schools that had a New Leader and that contribute to the achievement analysis are given by the "At least one year" row.

[8] Years of tenure are based on the maximum number of years a New Leader was at a particular school. Generally, this is the years a school has had a New Leader (as of 2008). For schools where the New Leader left before 2008, it refers to the years of tenure in the last year they were in the school. When a school had multiple New Leaders, the program year refers to the maximum year of tenure at a school.

[9] Because the Regents exam is not administered in a uniform grade, the NYC high school count refers to the number of program schools with any student achievement data by 2008. These counts include schools for which we could not produce separate school-level estimates due to small sample sizes.

Table 3.1.
Counts of New Leaders Schools by City

	K-8 Schools						
Principal Tenure	**Bal-timore**	**NYC**	**OUSD***	**Chi-cago**	**DC**	**Mem-phis**	**Total**
At least one year	18	33	17	40	33	19	160
At least two years	5	24	13	27	26	9	104
At least three years	0	15	6	14	10	2	47
At least four years	0	8	2	10	1	0	21
At least five years	0	1	0	2	0	0	3

* Counts for OUSD include charter schools in Oakland for which we have student-level data, but exclude other charter schools. Due to small sample size, schools with three or more years of tenure in Chicago are pooled, and schools with two or more years of tenure in DC are pooled.

4. METHODS

The goal of our analyses is to estimate the effect of the New Leaders program on student outcomes. Specifically, the program effect of interest is defined to be the difference between the outcome a student would have in a New Leaders school and the outcome that would occur had that student attended a school led by a non-New Leader. This is a standard way to conceptualize a program effect, but it does not capture all relevant effects of the New Leaders program. For instance, principals who did not go through the New Leaders training might nonetheless mimic the leadership strategies of the principals who did. This may be because districts that partner with New Leaders have "bought into" the vision of the organization and have adopted similar management practices. If "spillovers" such as this take place, then the effects we estimate would understate the total impact of the program. However, the program effect we hope to estimate would still be informative about the impact of attending a New Leaders school compared to attending a non-program school in the same district.[10]

A key aspect of this evaluation is to understand how the effect of the program changes as principals acquire experience as principals and specifically with their school. Since New Leaders typically come in with relatively little experience in school administration positions, their "learning curve" may be steeper than it is for other principals in comparable schools. Thus, we devote considerable effort toward unpacking how the effects evolve as principals get more experience.

The ideal way to estimate the impact of the New Leaders program would be to randomly assign principals to schools, and compare the outcomes of students attending the experimental schools to the non-New Leaders schools. Randomization would ensure that other factors that

[10] Estimating the total effect of the program that includes any spillover effects in schools led by non-New Leaders would only be possible with variation in the presence of any New Leaders principals. For instance, we could compare achievement in a district before and after the program was implemented. The clear limitation of such an analysis would be the inability to distinguish the program effect from general trends or the influence of other programs.

could affect student achievement are "balanced" between the program and control schools, which would imply that simple comparisons by New Leaders status would identify the impact of the program.

However, principal assignment is governed by a variety of district practices, regulations, and union agreements that vary across New Leader partner districts. New principals go to work in schools that have a principal vacancy. Principal vacancies may exist because a school is a new school, because the previous principal has left, or because the prior principal has been removed – either for poor performance or transferred to another school. Schools with a principal vacancy are not likely to be a random sample of all district schools. In some districts, such as Chicago, local school councils are able to choose the principal from the set of district-approved candidates (which includes the New Leaders candidates), and the district has little control over specific assignments. In other districts, the district determines assignments with little input at the school level. New Leaders works with districts to ensure that New Leaders candidates are considered for suitable available positions, but random assignment is not possible. Thus, as designed, the program does not lend itself to evaluations based on random assignment. More generally, this reasoning underscores that there are serious barriers to implementing any program evaluations based on principal random assignment.

Furthermore, New Leaders are often placed in schools that look very different from the typical school in a given district. Because many of these differences are along dimensions that are known to be highly correlated with achievement (such as socioeconomic status), estimates of the program effects may be misleading if they do not account for these confounding factors. There are two main sources of bias. One arises from students sorting across schools, and would occur if students attending New Leaders schools are higher (or lower) ability than the typical district student. The second is New Leaders sorting to certain schools. For instance, New Leaders are more likely to be in charter schools and startup schools than is typical in the cities we examine. To address the first type of bias, we estimate models that control for time-invariant student-level heterogeneity and control for time-varying observable characteristics. To address the second source of bias, we control for

observable school-level factors. In the remainder of this chapter, we describe the analytic approach in greater detail.

Although we believe that these methods are appropriate given the institutional context, it is important to recognize that in any evaluation such as this, there is always the possibility that the estimates are biased by omitted school or student characteristics. For instance, it may be that New Leaders are placed in schools with lower than average teacher quality. Alternatively, they may be placed in schools where students had unusually low test scores, which helped lead to the dismissal of the previous principal, and that test scores subsequently increase due to mean reversion rather than because of the practices of the New Leaders principal. Without experimental variation (or a clean "natural experiment"), the presence of these types of biases will always be possible no matter what statistical approach is used.

Our primary analytic approach uses longitudinal student-level data to estimate "value-added" models.[11] The goal of all value-added models is to isolate the incremental contribution of some intervention, in this case, the New Leaders for New Schools program, on student outcomes (McCaffrey et al., 2003). Because this approach exploits "interruptions" in the student's exposure to a New Leaders program, it can also be characterized as a quasi-experimental design known as interrupted "time-series" (with a control group), with data on "control" students who never receive exposure to the program.[12] We estimate models of the form:

$$(1)\quad Y_{ist} = \theta D_{st} + X_{st}\beta + W_{ist}\lambda + \alpha_i + \eta_t + \varepsilon_{ist}$$

where Y_{ist} denotes student achievement for student *i* in year *t* in school *s*, X_{st} is a vector of observed school-level covariates (such as

[11] An alternative approach would use longitudinal data at the school level and estimate models that control for school fixed-effects. This approach is feasible, but with aggregate data it is not possible to control for individual-level student characteristics or to control for unobserved student-level permanent heterogeneity. Furthermore, as discussed below, controlling for school fixed-effects is problematic given that many of the program schools are startups, which means that there is not within-school variation in treatment status for many schools.

[12] Because the non-New Leaders school constitute a form of a control group, we are essentially estimating difference-in-differences models.

principal experience), η_t is a year fixed-effect, W_{ist} is a vector of observed student-level covariates that may be time-varying (such as grade) or permanent (such as race), α_i is a student-specific intercept, and ε_{ist} is a random disturbance term. The key variable for this study is D_{st}, which denotes the program status of school *s* in year *t*. Because the effect of the program is theorized to change as New Leaders acquire experience, D_{st} is parameterized using an indicator for a New Leader in her first year, an indicator for a New Leader in her second year, and a New Leader in her third or higher year.[13]

A key modeling decision concerns α_i. The "fixed-effects" approach includes it as a parameter to be estimated (or "differenced out"). Fixed-effects models address the bias associated with differences between students in New Leaders and non-New Leaders schools by explicitly controlling for all time-invariant ("fixed") differences between students. Effectively, each student serves as his own control. Fixed-effects models exploit within-student changes in the exposure to the New Leaders "treatment." Transitions into or out of program status occur when students change schools (e.g., from elementary to middle school) or when the school a student attends gains or loses a New Leaders principal. In addition to accounting for fixed student-level factors, the models we estimate also include controls for time-varying student characteristics (such as age) and school characteristics (such as charter school status).[14] The list of covariates included in the models for each district can be found in Appendix A, Table A.1.

---

[13] These variables are based on the number of years a New Leader has served at the current school (in a handful of cases, New Leaders move from one school to another and so could have more years as a principal than as a principal at a given school).

[14] Alternative models are also possible--for example, rather than specifying individual-specific intercepts, which allows for heterogeneity across students in the level of performance, one might instead control for individual-specific time trends, thus allowing the growth trajectory of performance to vary across students. Another possibility is to estimate models where the dependent variable is achievement growth. Such models involve alternative statistical approaches (such as estimation in differences) which we do not consider in this paper but plan to explore in future work.

The estimated program impact derived from fixed-effects models can be understood as reflecting the differential changes in outcomes, conditional on the included time-varying factors, of students whose program status changes compared to the achievement growth of students who remain in non-NLNS schools throughout the study period.

The main advantage of fixed-effects estimation is that it eliminates biases arising from persistent differences in the characteristics of students attending program and non-program schools (Harris and Sass, 2006; Wooldridge, 2001). The drawback of the fixed-effects approach is that it "throws out" variation across students, thereby using the data less efficiently than an estimation procedure which used between-student variation. A second drawback is that the estimates may be identified off of the program status transitions of relatively few students. This is especially pertinent in the current study because many New Leaders are placed in new or "startup" schools. When the startups are elementary schools where the New Leader was in place from the school's inception, few students will experience a change from the non-program to the program state (this is less of a concern for startup middle and high schools since students will experience changes in treatment status when they transition from elementary to middle school).[15,16]

The second type of model we use is a random-effects model.[17] Unlike fixed-effects models, these models do not explicitly control for

---

[15] This is also the reason we do not estimate models with school fixed-effects. In particular, the school fixed-effect would be collinear with treatment for startup schools where the only principal observed at the school is a New Leader.

[16] Because the model is parameterized in such a way that the treatment effect varies with the level of experience of the New Leader principal, students in startup elementary schools can contribute to the identification of changes in the effect of the program with respect to experience.

[17] The models we estimate are one-way (student) random-effects models. An alternative approach is to estimate mixed random-effects models that include a school intercept. We have experimented with this approach for some cities and found the results to be largely consistent with the one-way random-effects estimates. However, based on simulations we conducted, the standard errors produced by the mixed models appear to be too small, and lead to over-rejection of the "no effect" null hypothesis.

permanent student heterogeneity. Instead, student heterogeneity in achievement levels is modeled as a random term that is unrelated to attending a NLNS school or a comparison school. The implicit assumption is that students in NLNS and non-NLNS schools do not differ in ways that are associated with achievement, at least after accounting for observable differences in student, school, and principal characteristics.[18] The main advantage of this approach is that, when the underlying modeling assumptions are correct, it makes more efficient use of the data and the estimates tend to be more precise than those from fixed-effects models. In particular, the random-effects models might be preferable for cases where most students in a school do not have data from a non-program period (which would likely be the case in startup elementary schools).

The concern about random-effects models is that they deliver biased estimates if students in NLNS schools are "different" from students in comparison schools in ways that are related to test performance and that are not measured by variables included as controls in the models. Therefore, fixed-effects models are often seen as preferable to random-effects models from the standpoint of producing unbiased estimates (Halaby, 2004). However, a recent study (Lockwood and McCaffrey, 2007) indicates that when both random-effects and fixed-effects estimates are biased, random-effects models may actually outperform fixed-effects methods in terms of bias minimization. In practice, the fixed- and random-effects estimates presented below are quite similar.

Because the tests, institutions, and student characteristics differ so much across cities, the estimation is done on a city-by-city basis. For each city, the estimated standard errors on the effect sizes are adjusted for clustering at the school-year level. To produce aggregate estimates, we take a weighted average of the city-level estimates. The goal of this exercise is to produce an estimate that reflects the average impact for a representative New Leader principal. Hence, the

[18] The random-effects models include the same controls that are used in the fixed-effect models. In addition, we also control for time-invariant student characteristics such as special education status, race, and ethnicity (dummies for African-American, Hispanic, and in OUSD, Asian), gender, Limited English Proficient program, free/reduced lunch, and an indicator for being old for grade.

weights are based on the number of New Leaders in a city, which reflects the fact that a city-level estimate with relatively few New Leaders underlying it should be downweighted in the calculation of the aggregate effect.

5. RESULTS

In this chapter, we present the pooled estimates of the effect of the New Leaders program on student achievement. City-level estimates are reported in Appendix D. Because the number of treated schools in a given city is relatively small, these estimates tend to be imprecisely estimated, which is the main reason we focus on the national estimates in this paper.

Table 5.1 shows the pooled effect size estimates of the coefficients on the New Leaders program status indicators along with estimated standard errors. In bold below the effect size is the implied effect on the percentile ranking (assuming that standardized test scores follow a normal distribution) of a student whose counterfactual test score would place them at the median of the test score distribution.[19]

The results in Table 5.1 suggest that being in a school led by a New Leader is associated with improved academic achievement for principals with three or more years of experience (estimates for second year principals are also positive and are statistically significant for math). For principals with three or more years of experience, the fixed-effects estimates imply an effect size of about 0.04 and the random-effects estimates suggest an effect size of about 0.06.[20] Because the

[19] In other words, the implied percentile ranking is given by $\Phi\left(\hat{\theta}\right)-0.5$, where Φ is the c.d.f. of a standard normal.

[20] These effects should be thought of as cumulative effects. In particular, the positive estimate for principals in their third or higher year represent the total difference in average achievement that is realized after the program has been in place for three or more years, and does not represent the incremental program effect relative to the effect of a second-year principal. Moreover, since the estimate of a first-year principal is negative, the estimates suggest that New Leaders may improve more than the year-3+ effect would suggest (although from a policymakers perspective this improvement is less important than the absolute year-3+ effect). The cumulative effects capture both the effect of the principal acquiring more experience and students acquiring more years of exposure to the program principal. Since students in, for instance, a Year 3 school will differ in how long they have been in a program school (because of mobility and because younger students will have entered the school more recently), these effects should be thought of as a weighted average of the effects for students with a given number of years of exposure to the program.

fixed- and random-effects models both produce very similar estimates the choice of modeling the student-level intercept does not appear to have an important effect on the qualitative or quantitative findings. While these estimates may not seem especially large, they are in line with estimates of the effect of other principal characteristics found in two recent studies. For instance, in New York City, Clark, Martorell, and Rockoff (2009) find an effect size of about .04 for math and .03 for reading when comparing a principal with five years of experience to a rookie. Branch, Hanushek, and Rivkin's (2009) estimates for Texas point to even smaller effects (about 0.025). Note that since our models control for principal experience of all principals, the estimates do not reflect a pure experience effect, but rather suggest that New Leaders with three or more years of experience are associated with larger achievement gains than would be realized in a school with similar observable characteristics (including principal experience). However, as noted above, some of the principal tenure data we have collected is incomplete or has some errors. This could result in biased estimates.[21]

The estimates for principals in their second year are positive, but are smaller. The estimates for math suggest that achievement is about 0.02 standard deviations higher in schools led by a second year New Leader compared than it would be in a non-program school. The estimates for reading are smaller and only marginally statistically significant in the random-effects model. For first-year New Leaders, the results indicate there is a negative relationship between New Leaders in their first year and student achievement.[22] The magnitude of this association,

[21] One reason to believe any such bias may not be very serious comes from the analysis reported in Appendix C. We estimated models for Memphis excluding principal characteristics and models that included rich principal characteristics, and the results were largely the same across these specifications. The extent to which this conclusion is valid, however, depends on the quality of the principal data provided by Memphis. In ongoing work, we are collecting improved principal information to further guard against these potential biases.

[22] Ideally, this comparison would reflect a comparison to other first-year principals. And to the extent our principal tenure data is complete and error free, it will be. However as noted above, there are several problems with the principal tenure data we are using, so it may be that the negative effect in the first year may reflect a pure tenure effect.

though, is small (an effect size of about .02 to .03). Further, this may reflect inadequate controls for principal experience rather than a true negative effect since New Leaders have less experience than typical district principals.

Table 5.1.
Pooled Program Effect Estimates, Elementary and Middle School

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
NLNS Impact in 1st Year	-.0192**	-.0295**	-.0232**	-.0335**
	(.00494)	(.00380)	(.00696)	(.00643)
	-0.77	**-1.18**	**-0.93**	**-1.34**
NLNS Impact in 2nd Year	.0160*	.00464	.0243**	.0125†
	(.00643)	(.00496)	(.00816)	(.00669)
	0.64	**0.19**	**0.97**	**0.50**
NLNS Impact in 3rd+ Year	.0381**	.0418**	.0598**	.0637**
	(.00720)	(.00539)	(.0130)	(.00841)
	1.52	**1.67**	**2.39**	**2.54**

Note: Standard error in parentheses. Effect on percentile ranking in bold. "**", "*" and "†" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively. 137 schools contributed to the 2008 overall estimate, and 105 schools contributed to the 2007 overall estimate. In 2008, there are 50 Year 1 schools, 45 Year 2 schools, 23 Year 3 schools, 17 Year 4 schools, and two Year 5 schools. 160 schools have at least one year of tenure, 104 schools have at least two years of tenure, 47 schools have at least three years of tenure, 21 schools have at least four years of tenure, and three schools have at least five years of tenure.

These results are consistent with the organization's theory of action which stipulates that the full benefit of the training and program support is only realized after several years, by which time a principal will have had time to institute changes and reforms to a school. However, it is important to recognize that not all New Leaders are observed with three or more years of experience. In particular many New Leaders had only been in place for one or two years by the 2008 school year. Additionally, some principals left before their third year. Thus, another interpretation of the pattern in Table 5.1 is that the differences in the estimates by year of experience reflects differences in the composition or principals contributing to the estimates rather than changes in the effect as principals acquire experience. To examine this possibility, we re-estimated the models excluding New Leaders who had less than three years of experience by the 2008 school year. In this

sample, comparing estimates of the treatment effect by year of experience would not be confounded by compositional effects. The results in Table 5.2 are similar to those in Table 5.1 – the estimated effect sizes increase with principal experience. This suggests that it is principal experience and not cohort effects (i.e., that the early New Leaders were always associated with higher achievement, even in their early years) that are driving the patterns observed in Table 5.1.

Table 5.2.
Pooled Program Effect Estimates for New Leaders With at Least Three Years of Experience, Elementary and Middle Schools

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
Impact in 1st Year, at least 3 Years of Tenure	-.0235	-.00288	-.0127	.00445
	(.0161)	(.0118)	(.0258)	(.0222)
	-0.94	**-0.11**	**-0.51**	**0.18**
Impact in 2nd Year, at least 3 Years of Tenure	.0401*	.0311†	.0534*	.0432†
	(.0173)	(.0172)	(.0249)	(.0226)
	1.60	**1.24**	**2.13**	**1.72**
Impact in 3rd Year, at least 3 Years of Tenure	.0464*	.0551**	.0705*	.0795**
	(.0204)	(.0146)	(.0302)	(.0205)
	1.85	**2.20**	**2.81**	**3.17**

6. CONCLUSION

This paper discussed the methodological challenges involved in estimating the effect of the New Leaders for New Schools program principals on student achievement, our approach to addressing these challenges, and preliminary findings through the 2007-08 school year. The estimates suggest that there is a positive association between attending a New Leaders school led by a principal in the third (or higher) year, and some evidence of a positive relationship for second-year New Leaders as well. There is a small negative relationship with achievement for first-year New Leaders.

There are two important caveats that should be borne in mind. First, these results are from an ongoing evaluation. In the next several years, we will repeat and extend the analysis using information on principals who have been placed after the 2008 school year. As such, the results will undoubtedly change. Second, these estimates are not based on an experimental design. Although we use a quasi-experimental interrupted time-series (with control) design that helps control for unobserved student-level heterogeneity, there remains the possibility that differences between students in program and non-program schools could lead to biases.

Future work will focus on three areas. First, we will refine the analysis of how the effects of the program evolve as principals acquire experience and explore incorporating years of student exposure to a New Leader in the models. Second, we will attempt to acquire better data on principal experience and student demographics to help control for potential confounds. As noted earlier, our data on principal tenure is sometimes incomplete and the level of detail on student demographics could be enhanced by collecting information on factors such as public housing residence. Finally, we will attempt to collect student-level data for all of California so that we can examine New Leaders placed outside of Oakland.

APPENDIX A.
LIST OF COVARIATES INCLUDED IN THE MODELS BY CITY

Because the data availability varies across cities as well as the need to control for whether a school participates in a given program differs across the study sites, the set of covariates included in the models is not the same for each city. Table A.1 shows the covariates that are included in the models for all six study cities.

Table A.1.
Covariates Included in the Models, by City

	Fixed-Effects Models	Additional Covariates in Random-Effects Models*
Chicago	Grade effects; School year effects; School-level effects; School averages of: free/reduced lunch, ESL, special education, white, Hispanic, African-American, Asian, old for grade, male, mobile; Number of students; Regular School; Charter; Magnet; Startup by year of startup (2002-2008); Principal tenure in job, tenure in school, and tenure in CPS, age, degree, sex, and race	Free/reduced lunch; ESL; Special education; Male; Hispanic; Asian; African-American; Old for grade
New York City	Grade effects; School year effects; School-level effects; Age; School averages of: free lunch, reduced lunch, ESL, special education, white, Hispanic, African-American, old for grade, mobility; Number of students; Empowerment school; Charter; Startup in 2002-04; Startup in 2005-08; Single-year of principal tenure indicator variables for tenure=1,6; tenure missing	Free/reduced lunch; ESL; Special education; Immigrant; Male; Hispanic; African-American
Washington DC	Grade effects; School year effects; School-level effects; School averages of: free/reduced lunch, ESL, special education, white, Hispanic, African-American, Asian, old for grade; Number of students; Regular School; Charter; Startup by year (2005-2008); Principal tenure	Free/reduced lunch; ESL; Special education; Male; Hispanic; African-American; White; Asian
OUSD	Grade effects; School year effects; School-level effects; Age; School averages of: free/reduced lunch, ESL, special education, white, Hispanic, African-American, old for grade, mobility; Number of students; Small school; Charter; Startup in	Free/reduced lunch; ESL; Special education; Male; Hispanic; African-American

	Fixed-Effects Models	Additional Covariates in Random-Effects Models*
	2003-05; Startup in 2006-08; Single-year of principal tenure indicator variables for tenure=1,6; Tenure missing	
Memphis	Grade effects; School year effects; School-level effects; School averages of: free/reduced lunch, ESL, special education, mobile, male, white, Hispanic, African-American, Asian, old for grade; Number of students; Regular School; Charter; Magnet; Startup by year (2004-2008); Principal tenure in job, tenure in school, and tenure in MCS, age, degree, sex, race, and number of tenures	Free/reduced lunch; ESL; Special education; Male; Hispanic; African-American; White; Asian; Mobile
Baltimore	Grade effects; School year effects; School-level effects; Age; School averages of: free/reduced lunch, ESL, special education, white, Hispanic, African-American, mobility; Number of students; Principal tenure; Tenure missing	Free/reduced lunch; ESL; Special education; Male; Hispanic; African-American

APPENDIX B.
INFORMATION ON PRINCIPALS

As noted in the text, New Leaders principals have less experience than typical district principals. Because principals may become more effective as they acquire experience, it is crucial that we control for these differences as thoroughly as possible. This appendix describes the principal tenure data we have collected for each of the study sites.

- CPS: We have consolidated information from separate tenure files received from CPS covering principals in the 2005-06 and 2006-07 school years and principals current as of March 2009. Tenure variables include tenure in the current principal position, tenure within the school, and tenure within CPS. We are able to infer tenure in earlier years for principals in place as of the 2005-06 school year, but we do not observe tenure for principals present in 2007-08 but not in 2008-09. Data for charters were unavailable.[23]
- NYC: We received updated data on NYC principal tenure that contains information on the number of years spent as a NYC district principal. Combined with data collected in the Year 1 and Year 2 analyses, we have complete data for all study years. The data appear to be reliable, but do not cover charter schools. Charter schools form a small portion of the overall NYC school population, but represent an important component of the New Leaders placements in NYC.
- Memphis: We received updated principal tenure data that corrected serious errors that we discovered were part of the data provided to us in earlier years. The data we received for the Year 3 study appears to be reliable and includes information on years of tenure within the district in any

[23] Charter school principals typically have less experience than principals in regular district schools (Gates et al., 2003). Our controls for charter status would account for these general differences. However we currently do not have any way to account for differences in tenure in charter schools between New Leaders principals and charter comparison schools.

position, years of tenure as a principal, years of tenure in the current principal position, and number of separate tenures as a principal. Because the principal data look to be of high quality and include a number of characteristics not available in other districts, we used Memphis as a "test case" to see how the estimates change when we omit principal controls from the model to help gauge whether there are likely to be important biases from having lower-quality data in other districts (see Appendix C).

- Baltimore: Principal tenure data was provided for principals in place in the 2007-08 school year. With this information, we can identify the year a principal serving in 2007-08 was hired. There are two noteworthy issues with the quality of this data. First, it does not provide any information on tenure in years prior to 2007-08 if there is a new principal in 2007-08. This is arguably less of a concern in Baltimore since New Leaders were only first placed there in the 2007 school year, and most were placed in the 2007-08 school year. Second, tenure data is missing for about one-third of the schools.
- Washington DC: We know tenure for principals who began in their schools after 1997, which includes a substantial majority of public school principals in the study. Our data are limited to principals in non-charter schools.
- OUSD: We have complete principal tenure data starting with the 2003-04 school year, with information on a principal's tenure at their current school. Prior to that, we only know tenure for principals who did not leave their school before the 2003-04 school year.

APPENDIX C.
ANALYSIS OF THE IMPORTANCE OF PRINCIPAL CONTROLS

One concern in comparing New Leaders principals to other principals is that New Leaders principals are disproportionately first-time principals that have a different career path than typical district principals. If principal effectiveness increases over time as principals gain experience, New Leaders principals might appear less successful simply due to inexperience. More generally, failure to account for any characteristic of principals that independently impacts student achievement and is correlated with New Leaders program assignment might bias estimates of the impact of the program.[24] As described in the preceding appendix, we have collected data from the individual districts to obtain principal biographical data that can be used to control for factors such as principal experience in the analysis and we continue to work with partner districts to improve available data.

This appendix examines the extent to which the findings in the text are affected by the inclusion of controls for principal characteristics. Memphis, which provided detailed information regarding its principals, provides a useful laboratory for examining this question. Table C.1 reports regressions estimates of the effect of the program on test scores; for simplicity the effects of the program are captured using a single program status indicator rather than breaking out the effects by year of tenure as we do in the main specifications reported in the text. Separate estimates for math and reading and random- and fixed-effects models are presented. Column I reports effect estimates that control for student and school characteristics but no principal characteristics. Column II expands the controls to include linear and quadratic terms in years of experience as a principal. Column III adds linear and quadratic terms for principal age, tenure within MCS in any position, and tenure

[24] In theory, New Leaders principals might perform differently not because of direct impacts of New Leaders training, but because the screening process for New Leaders identifies individuals with innate characteristics that make them more or less effective as principals. The extent to which achievement differences that arise from such sorting should be characterized as program effects remains an open question.

in the current position. Column IV adds additional flexibility to this specification be including separate dummy variables for five-year intervals of principal tenure, MCS tenure, and tenure in current position as opposed to simple time trends. The use of dummy variables allows the effects of principal tenure to vary non-monotonically over time. Column V adds further fixed effects that control for principal race, gender, educational attainment, and number of separate tenures as principal.

Table C.1.
Impact of Varying Principal Controls on Program Effect Estimates for Memphis, Lower Grades

	I	II	III	IV	V
Fixed-effects					
Impact on math scores	-.0179	-.0130	-.00930	-.0126	.00147
	(.0263)	(.0466)	(.0464)	(.0238)	(.0297)
	-0.71	-0.52	-0.37	-0.50	0.06
Impact on reading scores	-.0250	-.0282	-.0168	-.0115	.00410
	(.0155)	(.0185)	(.0195)	(.0152)	(.0159)
	-1.00	-1.12	-0.67	-0.46	0.16
Random-effects					
Impact on math scores	-.0366	-.0367	-.0386	-.0361	-.0297
	(.0330)	(.0486)	(.0466)	(.0303)	(.0360)
	-1.46	-1.46	-1.54	-1.44	-1.18
Impact on reading scores	-.0458†	-.0544**	-.0499*	-.0369	-.0264
	(.0244)	(.0197)	(.0196)	(.0232)	(.0266)
	-1.83	-2.17	-1.99	-1.47	-1.05
Quadratic in principal tenure?	No	Yes	Yes	No	No
Quadratics in additional tenure variables?	No	No	Yes	No	No
Fixed effects for all tenure variables?	No	No	No	Yes	Yes
Control for additional principal characteristics?	No	No	No	No	Yes

For the random-effects model, the estimated program effects change little depending on whether or how principal characteristics are controlled. For fixed-effects models, there is a larger difference between models with and without demographic controls. However, in all

cases, the point estimates with numerous and flexible controls for principal characteristics are well within the confidence intervals of the estimates that include no principal controls, suggesting that the use of alternative methods for accounting for principal characteristics is unlikely to appreciably impact the findings of the student achievement analysis.

APPENDIX D.
CITY-LEVEL ESTIMATES

This appendix reports the estimates from each of the six cities included in the analysis. The estimates reported in the text are a weighted average of the city-level estimates.

Table D.1 shows results for Oakland. The results point to positive effects on reading achievement for students in schools led by New Leaders with 3+ years of experience. The estimates for math are also positive, but not precisely estimated. Estimates for first-year New Leaders are negative but small in magnitude and not statistically significant.

Table D.1.
Student Achievement Effect Sizes for OUSD

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
NLNS Impact in 1st Year, all schools	-.0438	-.0366	-.0237	-.0272
	(.0348)	(.0251)	(.0418)	(.0366)
	-1.75	-1.46	-0.95	-1.09
NLNS Impact in 2nd Year, all schools	.0405	.0434	.115	.112†
	(.0514)	(.0471)	(.0727)	(.0614)
	1.62	1.73	4.59	4.45
NLNS Impact in 3rd+ Year, all schools	.128†	.224**	.179	.278**
	(.0698)	(.0427)	(.114)	(.0542)
	5.11	8.87	7.12	10.96

Note: Standard error in parentheses below effect size (adjusted for clustering at the school-year level). Effect on percentile ranking in bold. "**", "*", "+" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively.

Table D.2 shows results for Washington, D.C. Estimated effects for 3rd year principals are again positive, this time for both math and reading. On the other hand, the estimates for first- and second-year New Leaders are small and statistically insignificant.

Table D.2.
Student Achievement Effect Sizes for Washington DC

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
NLNS Impact in 1st Year, all schools	.00949	-.0427*	.0265	-.0164
	(.0219)	(.0203)	(.0425)	(.0406)
	0.38	**-1.70**	**1.06**	**-0.65**
NLNS Impact in 2nd Year, all schools	.0557†	5.89E-4	.0811	.0320
	(.0299)	(.0255)	(.0495)	(.0437)
	2.22	**0.02**	**3.23**	**1.28**
NLNS Impact in 3rd+ Year, all schools	.0937*	.0570†	.166**	.137*
	(.0405)	(.0343)	(.0494)	(.0588)
	3.73	**2.27**	**6.59**	**5.43**

Note: Standard error in parentheses below effect size (adjusted for clustering at the school-year level). Effect on percentile ranking in bold. "**", "*", "+" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively.

Table D.3.
Student Achievement Effect Sizes for Memphis

	Fixed Effects		Random Effects	
	Math	**Reading**	**Math**	**Reading**
Impact in 1st Year, all schools	.0148	-.00112	-.0151	-.0322
	(.0309)	(.0179)	(.0364)	(.0317)
	0.59	**-0.04**	**-0.60**	**-1.28**
Impact in 2nd Year, all schools	-.0325	.0169	-.0708	-.0225
	(.0474)	(.0198)	(.0548)	(.0288)
	-1.30	**0.67**	**-2.82**	**-0.90**
Impact in 3rd Year, all schools	.0453	.0792	-.0320	-.00289
	(.0446)	(.0550)	(.0379)	(.0595)
	1.81	**3.16**	**-1.28**	**-0.12**

Note: Standard error in parentheses below effect size (adjusted for clustering at the school-year level). Effect on percentile ranking in bold. "**", "*", "+" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively.

Estimates for Memphis are shown in Table D.3. Estimates are somewhat larger in the fixed-effects specification but in all cases the estimates are not statistically significant.

Table D.4 shows results for Chicago. Estimates for first-year principals are negative and statistically significant. The effects for principals with two or more years of experience are small and not statistically significant.

Table D.4.
Student Achievement Effect Sizes for Chicago

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
NLNS Impact in 1st Year, all schools	-.0608**	-.0367**	-.0655**	-.0437*
	(.0132)	(.0126)	(.0218)	(.0209)
	-2.43	**-1.46**	**-2.61**	**-1.74**
NLNS Impact in 2nd Year, all schools	.0138	-.00235	.00851	-.0139
	(.0190)	(.0179)	(.0256)	(.0239)
	0.55	**-0.09**	**0.34**	**-0.55**
NLNS Impact in 3rd+ Year, all schools	.0262	.0167	.0255	.0110
	(.0233)	(.0163)	(.0432)	(.0269)
	1.04	**0.67**	**1.02**	**0.44**

Note: Standard error in parentheses below effect size (adjusted for clustering at the school-year level). Effect on percentile ranking in bold. "**", "*", "+" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively.

The results for New York City in Table D.5 are similar to those for Chicago. Impacts for first-year principals are negative and statistically significant. For second-year and higher, the estimates are also mainly negative, but are smaller in magnitude and not statistically significant.

Finally in Table D.6, the results for Baltimore do not show statistically significant positive or negative effects and the magnitudes of the effect sizes are relatively small.

Table D.5.
Student Achievement Effect Sizes for New York City

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
NLNS Impact in 1st Year, all schools	-.0327	-.0320**	-.0529*	-.0480**
	(.0223)	(.00976)	(.0268)	(.0147)
	-1.30	**-1.28**	**-2.11**	**-1.91**
NLNS Impact in 2nd Year, all schools	-.0230	-.0237	-.0361	-.0342
	(.0199)	(.0256)	(.0220)	(.0280)
	-0.92	**-0.94**	**-1.44**	**-1.36**
NLNS Impact in 3rd+ Year, all schools	-.0249	-.0230	-.0145	-.0126
	(.0348)	(.0196)	(.0418)	(.0222)
	-0.99	**-0.92**	**-0.58**	**-0.50**

Note: Standard error in parentheses below effect size (adjusted for clustering at the school-year level). Effect on percentile ranking in bold. "**", "*", "+" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively.

Table D.6.
Student Achievement Effect Sizes for Baltimore

	Fixed-Effects		Random-Effects	
	Math	**Reading**	**Math**	**Reading**
Impact in 1st Year, all schools	.0326	-.00773	.0257	-.0227
	(.0328)	(.0236)	(.0540)	(.0381)
	1.30	**-0.31**	**1.03**	**-0.90**
Impact in 2nd Year, all schools	.0331	.0765†	.0382	.0818
	(.0556)	(.0394)	(.0869)	(.0556)
	1.32	**3.05**	**1.53**	**3.26**

Note: Standard error in parentheses below effect size (adjusted for clustering at the school-year level). Effect on percentile ranking in bold. "**", "*", "+" indicate statistical significance at the 1%, 5%, and 10% significance level, respectively.

APPENDIX E.
DESCRIPTIVE STATISTICS BY CITY

This appendix presents school-level descriptive statistics by city, and by New Leaders program status. The results pertain to data from the 2007-08 school year.[25]

WASHINGTON DC

Table E.1 shows descriptive statistics for Washington DC schools. Schools with New Leaders look quite similar to other district schools, with the exception that a large number are charter schools. Program schools also have slightly fewer special education students.

Table E.1
Descriptive Statistics for Washington DC, by Program Status

	Comparison	Year 1	Year 2+
African-American	0.83	0.83	0.88
Hispanic	0.11	0.14	0.07
Asian	0.01	0	0.01
Boy	0.52	0.5	0.49
Free/Reduced Lunch	0.68	0.71	0.74
Special Ed.	0.16	0.09	0.12
ESL Participant	0.1	0.12	0.09
Top Half Math	0.43	0.45	0.4
Top Half Reading	0.43	0.46	0.43
Number of Students	212.78	311.56	200.64
Principal Tenure	2.83	1	1.56
Charter	0.29	0.78*	0
Startup	0.08	0.11	0
Number of Schools	161	9	19

* Difference relative to comparison is statistically significant at 10% level.

MEMPHIS

Table E.2 shows that many students in Memphis come from disadvantaged backgrounds, with over 85 percent of students in grades 1-8 receiving free/reduced price lunch. Comparing New Leaders to non-New

[25] Schools that are no longer New Leaders in 2008 are not considered program schools.

Leaders schools, the clearest difference is with principal tenure and tenure in other positions in MCS.

Table E.2
Descriptive Statistics for Memphis, by Program Status

	Com-parison	Year 1	Year 2+
African-American	88%	87%	96%
Hispanic	5%	8%	2%
Asian	1%	1%	0%
Boy	51%	52%	55%
Free/Reduced Lunch	86%	89%	93%
Special Ed.	14%	13%	14%
ESL Participant	5%	7%	2%
Top Half Math	48%	45%	44%
Top Half Reading	48%	43%	45%
Number of Students	484.09	603.5	293.3
Charter	6%	0%	0%
Magnet	16%	10%	17%
Startup	2%	20%	0%
Tenure with MCS, any Position	20.9	9.1	9.5
MCS Principal Tenure	6.2	1.5	2.4
Tenure at School	4.9	1.1	2.4
Number of Schools	137	10	6

* Difference relative to comparison is statistically significant at 10% level.

OAKLAND

As seen in Table E.3, Year 1 program schools in OUSD tend to have a higher fraction of Hispanic students, ESL participants and lower levels of achievement. However, since there are few Year 1 program schools, these differences are not statistically significant. Year 2+ program schools also have more Hispanic and ESL students, and are substantially more likely to be part of the Small Schools program or be startup schools.

Table E.3
Descriptive Statistics for OUSD, by Program Status

	Comparison	Year 1	Year2+
African-American	40%	38%	29%
Hispanic	34%	43%	62%*
Other Ethnicity	18%	17%	10%
Reduced Lunch	63%	65%	82%*
Special Ed.	10%	5%	6%
ESL Participant	31%	41%	57%*
Top Half Math	45%	33%	45%
Top Half Reading	46%	38%	42%
Number of Students	327.57	323.33	281.23
Principal Tenure	3.74	2	3.54
Small School	18%	33%	69%*
Startup	30%	0%	69%*
Number of Schools	77	3	13

* Difference relative to comparison is statistically significant at 10% level. These results include two program charter schools in OUSD, but not any program schools in the Bay Area located outside of Oakland.

CHICAGO PUBLIC SCHOOLS

Descriptive statistics for Chicago are shown in Table E.4. There are a number of noteworthy differences between program and comparison schools. Program schools tend to have fewer Hispanics, more African-Americans, and fewer low-income students than comparison schools. Achievement in the program schools is also lower in program schools, especially in Year 2 and beyond schools. We also see statistically significant differences in principal experience, the number of students, and the fraction of schools that are charter schools.

Table E.4
Descriptive Statistics for Chicago, by Program Status

	Comparison	Year 1	Year 2+
African-American	56%	85%*	65%
Hispanic	31%	6%*	29%
Asian	3%	1%	1%
Boy	51%	52%	50%
Low Income	84%	68%*	84%
Mobile	26%	27%	19%*
Free/Reduced Price Lunch	32%	15%	39%
Special Ed.	15%	11%	12%
ESL Participant	13%	1.72%*	11%
Top Half Math	47%	35%*	46%
Top Half Reading	48%	44%	47%
Number of Students	634	361*	847*
Principal Experience (months)	52.06	12.1*	31.98
Years in CPS	21.19	7.11*	9.88*
Tenure at School (months)	55.39	11.79*	31.98*
Charter	6%	0%	28%*
Magnet	8%	0%	4%
Startup	1%	25%*	0%
Number of Schools	531	12	25

* Difference relative to comparison is statistically significant at 10% level.

NEW YORK CITY

Table E.5 shows that Year 1 program schools in New York City are much smaller than comparison schools, have principals with significantly less tenure, and are more likely to be charter schools or startups. Achievement levels are also lower than comparison schools, although these differences are not statistically significant. Year 2+ program schools draw on a more disadvantaged population than comparison schools. Students in these schools are more likely to be African-American, receive free or reduced-price lunch and have lower levels of achievement than comparison schools. These schools are also smaller (although not as small as the Year 1 schools), have principals with less tenure, and are more likely to be charters, Empowerment schools and startups than comparisons.

Table E.5
Descriptive Statistics for NYC, by Program Status

	Comparison	Year 1	Year 2+
African-American	34%	40%	55%*
Hispanic	38%	44%	37%
Free Lunch	61%	70%	74%*
Reduced Lunch	7%	12%	10%*
Special Ed.	10%	3%	5%
ESL Participant	11%	4%	9%
Top Half Math	46%	37%	36%*
Top Half Reading	46%	32%	36%*
Number of Students	526.75	170.33*	359.5*
Principal Tenure	4.59	0.26*	2.54*
Charter	4%	33%*	20%*
Empowerment	28%	33%	55%*
Startup	21%	100%*	80%*
Number of Schools	1170	6	23

* Difference relative to comparison is statistically significant at 10% level.

BALTIMORE

In contrast to some other districts, New Leaders in Baltimore are placed in schools that are not strikingly different from typical Baltimore schools (Table E.6). In fact, the new cohort of program schools may serve a relatively advantaged population, as evidenced by the lower fraction of African-American students and the fact that math scores are somewhat higher (although not statistically significantly so).

Table E.6
Descriptive Statistics for Baltimore Schools, by Program Status

	Comparison	Year 1	Year 2+
African-American	88%	72%*	94%
Hispanic	3%	6%	1%
Free/Reduced Lunch	72%	74%	69%
Special Ed.	17%	17%	13%
ESL Participant	2%	4%	0%
Top Half Math	50%	56%	54%
Top Half Reading	47%	46%	54%
Number of Students	424.19	404.55	499
Number of Schools	142	11	5

* Difference relative to comparison is statistically significant at 10% level.

REFERENCES

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin (2008). *Principal turnover and effectiveness*. Unpublished manuscript.

Brewer, D. 1993. "Principals and Student Outcomes: Evidence from U.S. High Schools." *Economics of Education Review* 12(4): 281-292.

Card, D. & Rothstein, J. (2006). *Racial Segregation and the Black-White Test Score Gap.* National Bureau of Economic Research: Working Paper No. 12078.

Clark, D., P. Martorell, and J. Rockoff (2009). *School principals and school performance*. Unpublished manuscript.

Coelli, Michael, and David Green (2009). *Leadership effects: School principals and student outcomes*. Unpublished manuscript.

Gates, S., J. Ringel, L. Santibanez, K. Ross, and C. Chung (2003). *Who is leading our schools?: An overview of school administrators and their careers*. Santa Monica, CA: RAND Corporation, MR-1679-EDU.

Halaby, C. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, 30: 507-544.

Harris, D. N., and T. R. Sass (2007). *The effects of NBPTS-certified teachers on student achievement*. National Center for Analysis of Longitudinal Data in Education Research, Working Paper No. 4.

Heckman, J. (1979) Sample selection bias as a specification error. *Econometrica,* 47, 153–61

Leithwood, K., K. S. Louis, S. Anderson, and K. Wahlstrom (2004). *Review of research: How leadership influences student learning*. Minneapolis, MN: Center for Applied Research and Educational Improvement, University of Minnesota.

Lockwood, J. R., and D. McCaffrey (2007). *Controlling for student heterogeneity in longitudinal achievement models*. Santa Monica, CA: RAND Corporation, RP-1266-EDU.

McCaffrey, D., D. Koretz, J. R. Lockwood, and L. Hamilton (2004). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation, MG-158-EDU.

New Leaders for New Schools (2009). *Principal Effectiveness: A New Principalship to Drive Student Achievement, Teacher Effectiveness, and School Turnarounds with Key Insights from the UEF™*. New York, NY: New Leaders for New Schools. Available on-line: http://www.nlns.org/uef.jsp

Teske, P. and M. Schneider. 1999. "The Importance of Leadership: The Role of School Principals." http://www.businessofgovernment.org/pdfs/Import_of_Leadership.pdf

Waters, T., R. J. Marzano, and B. McNulty (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.

Wooldridge, J. (2002). *Econometric analysis of cross sections and panel data*. Cambridge, MA: MIT Press.