MATHEMATICA
Policy Research

REPORT

Impacts of New Leaders on Student Achievement in Oakland

May 19, 2014

Kevin Booker

Jaime Thomas

CONTENTS

TABLES

EXECUTIVE SUMMARY

This report presents our examination of impacts of principals trained by New Leaders on student achievement in Oakland Unified School District (OUSD). For the last 10 years, New Leaders has partnered with OUSD in efforts to improve the recruitment and training of effective principals and school leaders. A recent evaluation report by RAND (Gates et al. 2014) described New Leaders' programs across the country, and this report complements that work by taking a closer look at New Leaders' work in Oakland. These two reports fill an important gap in the literature, as little research has been done looking at the impacts of principal training programs on student achievement.

Using student-level data from 2006–2007 to 2012–2013, we compare the performance of students who attended New Leaders-led schools in OUSD with the performance of other OUSD students. This comparison accounts for differences in the characteristics and baseline achievement levels of students attending different schools. We find positive and statistically significant effects of attending a New Leaders-led school on student achievement in math and English language arts (ELA). These effects correspond to approximately four months of additional learning in math over a three-year period, and one and one-half months of additional learning in ELA.

The report also presents our examination of the performance of New Leaders-led OUSD schools compared to schools in other California (CA) districts. A comparison with schools outside of OUSD is important because some New Leaders activities, such as principal selection methods and influencing the use of data driven instruction, could have spillover effects that impact all OUSD schools. To make this comparison, the report includes a school-level analysis comparing changes in average test scores between 2003 and 2013 for New Leaders-led schools in OUSD, relative to a matched comparison sample of non-New Leaders schools statewide. We find that from 2003 to 2013 New Leaders-led schools in OUSD had greater increases in average test scores in both ELA and math than either other OUSD schools or schools in the matched statewide comparison sample. The results of the school-level analysis are consistent with the student-level results, and suggest that the positive results would likely hold in comparisons with a representative statewide comparison group.

I. INTRODUCTION

This report describes the examination of impacts of principals trained by New Leaders on student achievement in Oakland Unified School District (OUSD). Founded in 2000, New Leaders has trained more than 1000 school leaders who are raising student achievement and graduation rates in high-need schools across the country. The New Leaders' selection process is highly competitive, with substantial outreach to identify high quality applicants. New Leaders supports aspiring principals through a structured residency program and ongoing training and support after they assume leadership positions in schools. The New Leaders' mission is to ensure high achievement for all children, especially children of color growing up in poverty. This report finds positive and statistically significant effects of attending a New Leaders-led school in OUSD on student achievement.

A recent evaluation report by RAND (Gates et al. 2014) described the examination of the impacts of New Leaders' principal training on student achievement. One point the RAND report makes is that a helpful next step would be to compare achievement for New Leaders' students with students in schools outside of the district, because of the risk of spillover effects throughout the district from influence on activities such as principal selection and the use of data driven instruction, even at schools that are not led by a New Leaders principal. This study is an attempt to fill this need for additional research making comparisons with students in external comparison districts.

New Leaders funded this evaluation to examine the effects of New Leaders' efforts to improve student achievement in the San Francisco Bay Area, drawing on New Leaders' long-standing relationship with OUSD. For the past 10 years, New Leaders has partnered with OUSD and Bay Area charter schools to develop and implement improved processes for selecting, training, and supporting effective principals and teacher leaders.

To learn about how New Leaders' processes have improved student achievement in OUSD, this evaluation addresses the following research questions:

1. How are students who have attended New Leaders-led schools performing relative to other students in OUSD?

2. How are New Leaders-led schools in OUSD performing relative to similar schools and schools statewide?

We address these research questions by conducting student- and school-level analyses. In the student-level analyses, we use data from OUSD to compare student achievement and achievement growth for different groups of students. We consider students within OUSD who had a New Leader for different lengths of time and compare results for these students to students within OUSD who never had a New Leader. We find that students who attended a New Leaders-led school show greater achievement growth in ELA and math than other OUSD students.

Our original research design was to also include a student-level comparison with a sample of schools from external comparison districts, with the sample chosen to be closely matched with OUSD in terms of average initial achievement levels and demographics. We approached a number of districts that could provide such a comparison sample, but the only two comparison

districts that agreed to participate did not form a close match for OUSD schools, with significant differences in average student demographics and average baseline achievement levels, so we were unable to include this component of the research design in the report. More details on the comparison with these two external districts are available in the appendix.

In the school-level descriptive analyses, we compare average test scores and changes in average test scores for different groups of schools. We compare results for schools within OUSD that had a New Leader for different lengths of time to three different comparison groups: 1) non-New Leaders schools within OUSD, 2) a sample of non-OUSD schools matched using propensity score matching to be similar to OUSD schools, and 3) all non-OUSD schools statewide. The results of the school-level analyses are consistent with the student-level results, and suggest that the positive results for New Leaders from student-level comparisons within OUSD would likely hold in student-level comparisons with a representative statewide comparison group.

In this report, we discuss our data sources and analysis methods used to answer the research questions (Section II). We describe our results from the school- and student-level analyses in Section III. Section IV provides our conclusions.

II. DATA AND METHODS

A. Student-level analyses

   The goal of this study is to produce the best possible estimate of the average impact of having a New Leaders-trained principal on students' outcomes, compared to outcomes for similar students at other schools. Achieving this goal requires identifying an appropriate comparison sample of students who are similar to students in New Leaders schools, and collecting the necessary data to track individual students' outcomes over time.

   For the student-level analyses, we used data provided by New Leaders containing information on which OUSD schools had New Leaders, the period of time a school had a New Leader, and the role of the New Leader. The dependent variables in our analyses were English language arts (ELA) and math California Standardized Test (CST) scores for grades 2 through 8. These scores are standardized to z-scores by school year, grade, and subject using CA state means and standard deviations. A z-score represents the number of standard deviations above or below the CA mean test score in that subject, grade, and year. We used the following student characteristics as covariates: race/ethnicity (white, black or African American, Hispanic, Asian, or other) and participation in special programs (gifted education or special education). We used data from the 2006–07 to the 2012–13 school years. The sample sizes varied by analysis, with the OUSD sample containing approximately 50,000 unique students (80,000 student-year observations with current and prior-year test scores).

   We explored several different definitions of the intervention group, students who had a New Leader. Intervention group 1, our primary definition on the analysis, included students who attended a New Leaders-led school for at least two years, where a New Leaders-led school is defined as one that had a New Leader for at least three years. These restrictions grew out of New Leaders' experience that it takes at least three years for a new principal to have a substantial impact at a school, and that a student would need to be exposed to the school for at least two years to see a substantial impact in their outcomes compared to being at a school that never had a New Leader's principal. This definition was agreed on as the primary treatment definition prior to any data analysis. The main alternative definition is intervention group 2, which is an unrestricted sample containing any students who ever had a New Leader.

   The primary student-level analytic method is a quasi-experimental model that compares student achievement levels, controlling for prior achievement and other background characteristics. The student's current test scores in ELA and math are the outcome measures, and the model includes prior test scores in both subjects as controls for the student's prior achievement. This involved regressing the ELA or math test score from year t on an intervention group indicator, the ELA and math test scores from year t – 1, race/ethnicity indicators (with white as the omitted category), indicators for involvement in gifted or special education, school-average prior achievement, and grade-by-year indicators, as in equation (1). Standard errors were clustered at the school level.

$$(1) \quad \begin{aligned} test_{it} = {} & \beta_0 + \beta_1 NL_i + \beta_2 test_{i,t-1} + \beta_3 black_i + \beta_4 hispanic_i + \beta_5 asian_i + \beta_6 other_i \\ & + \beta_7 gifted_i + \beta_8 specialed_i + \beta_9 grade * year_{i,t} + \varepsilon_{i,t} \end{aligned}$$

The treatment group is defined by whether the students the school attended was led by a New Leaders-trained principal, and thus varies by which school each student attended in a given year. Thus students can move in and out of treatment status over time as they switch schools, or as leadership changes at the school they attend.

A key consideration is identifying the most appropriate comparison group to use in comparing student achievement growth across groups of schools. Ideally such a comparison group would consist of schools similar to those in the New Leaders treatment group based on observable characteristics, such as average initial achievement levels and student demographics. The best way to achieve such a comparison group would be to recruit several large and diverse school districts to provide comparison schools, then use propensity score matching to select schools from among that pool a closely matched comparison sample. This was the original research design for this study. Unfortunately, there are only a handful of districts large enough and adequately demographically comparable to furnish such a well-matched pool of comparison schools, and all of them declined to participate in the study. The two districts that did agree to participate, which were large enough to serve as a comparison pool, are not a very close match with the New Leaders-led schools. See the appendix for more details about these two districts and the results from an analysis including them as the comparison group.

Another possibility is to make comparisons within Oakland USD, comparing student-achievement growth for students attending New Leaders-led OUSD schools with growth for students at other OUSD schools. There are several reasons to look at a within-district comparison. One is that the average student characteristics are very similar between New Leaders-led OUSD schools and other schools in OUSD, with close matches in both average initial test score levels and student demographics (see Table II.1). Another is that by being in the same school district, these schools share a common history through this time period, and are less likely to vary based on unobserved variation in district policies. The schools within OUSD may also be more similar in terms of unobserved characteristics such as their teacher populations and recruiting pools than schools in other districts would be. The primary concern with making comparisons within OUSD alone is that the district worked quite closely with New Leaders in a number of areas, some of which could have had systemic effects across the entire system of training principals. Examples of New Leaders work that could lead to spillover effects include adoption of principal selection methods and influence on the use of interim assessments and data driven instruction. This potential for spillover effects means that whatever impacts are estimated for New Leaders-led schools may be understating the full impacts once spillover effects are taken into account.

Table II.1. Comparison of key characteristics of New Leaders-led schools and other OUSD schools

Characteristic	New Leaders-led schools (3+ years)		OUSD		
	Mean	SD	Mean	SD	*p*-Value
Average ELA test score, 2003	-0.49	1.04	-0.62	1.00	0.658
Average math test score, 2003	-0.20	0.88	-0.51	0.80	0.176
Title I eligibility	0.88	0.33	0.86	0.35	0.825
student-teacher ratio	18.08	2.27	20.67	17.96	0.556
Percentage white	0.09	0.16	0.08	0.18	0.852
Percentage black	0.37	0.21	0.48	0.29	0.137
Percentage Hispanic	0.33	0.26	0.32	0.27	0.962

Source: California Department of Education STAR Data, 2003; National Center for Education Statistics Common Core of Data, 2003; New Leaders data

Note: Elementary and middle schools are included. Test scores are standardized using the means and standard deviations from the statewide sample for the relevant year. The reported *p*-values are from tests of the statistical significance of the difference in each measure between schools led by a New Leader for at least three years and each comparison group.

OUSD = Oakland Unified School District; ELA = English language arts

B. School-level analyses

Although the student-level analysis provides the best estimates of impacts, through its ability to examine achievement for individual students over time, it has some key limitations. The biggest limitation is that student-level data were unavailable for most districts, and only available in OUSD for a limited number of years, with the earliest being 2006–07. Although school-level data do not provide the ability to control for prior achievement at the individual students level, the fact that school-level data are available statewide going back to 2002–03 allows us to use the school-level data to help understand the results we find at the student-level, and to try to predict what might be found with additional student-level data.

The school-level data used in this study consisted of publicly available school-level data from the CA Department of Education (CDE) and the National Center for Education Statistics Common Core of Data (CCD) for the school-level analysis. From the CDE, we obtained school-level average ELA and math CST scores from the 2002–03, 2006–07, and 2012–13 school years for grades two through eight, standardized using state-level means and standard deviations for each year, grade, and subject. From the CCD, we obtained the following school characteristics for the 2002–03 school year: Title I eligibility, student-teacher ratio, percentage of these groups: white, black or African American, and Hispanic.

For the school-level analysis, we defined the intervention group to be elementary and middle schools within OUSD that had a New Leader for at least three years. To compare test scores and test score growth in New Leaders-led schools with other schools throughout the state, we considered several comparison groups. Comparison group 1 contained OUSD schools that never

had a New Leader. Comparison group 2 contained a sample of schools matched to OUSD schools. Comparison group 3 contained schools statewide, including schools within OUSD that never had a New Leader. Table II.2 contains the sample sizes for each of these groups.

To create the matched comparison sample of schools, we used a propensity score-matching algorithm to select schools similar to OUSD schools on the following characteristics from the 2002–2003 school year: average ELA and math test scores, Title I eligibility, student-teacher ratio, and percentage white, black or African American, and Hispanic. The matched comparison sample contained 79 schools. Nineteen of these schools were in the Los Angeles Unified School District and 12 were in the San Francisco Unified School District. Other districts contained, at most, four schools each (Table II.2). In 2002–2003, the New Leaders-led schools were similar in average achievement levels to other OUSD schools and to the matched comparison schools.

The school-level analysis examined changes in average test scores at the school-level, comparing these between New Leaders schools and the different comparison groups. These comparisons take place over three time periods: 2002–2003 to 2006–2007, 2006–2007 to 2012–2013, and 2003–2003 to 2012–2013. Note that because treatment status is defined by having a New Leaders principal at any time during the relevant period, some schools included in the treatment group may not be led by a New Leaders principal during the final year of that period, when the student outcomes are measured.

Table II.2. Intervention and comparison group sample sizes, school-level analysis

	Number of schools
OUSD	
Intervention group schools (3+ years with a New Leader)	17
Other New Leaders schools (1–2 years with a New Leader)	7
Non-New Leaders schools (0 years with a New Leader)	65
Total number of schools, OUSD	**89**
Districts containing matched comparison schools	
Los Angeles Unified	19
San Francisco Unified	12
San Diego Unified	4
Fresno Unified	3
Sacramento City Unified	3
Vallejo City Unified	3
Compton Unified	2
Emery Unified	2
Hayward Unified	2
Inglewood Unified	2
Ravenswood City Elementary	2
West Contra Costa Unified	2
Other Districts	23
Total number of schools, all other matched districts	**79**
Statewide	
Total number of schools, statewide	**6,785**

Source: California Department of Education STAR Data, 2003; National Center for Education Statistics Common Core of Data, 2003; New Leaders data.

Note: Elementary and middle schools are included. We matched OUSD schools to, at most, one comparison school within California, using data from 2003 on the following characteristics: school-level average ELA and math test scores, Title I eligibility, student-teacher ratio, and race/ethnicity distribution (percentage white, percentage black, and percentage Hispanic).

OUSD = Oakland Unified School District; ELA = English language arts

III. RESULTS

A. Student-level analyses

The primary research question for this study is whether schools led by a New Leaders-trained principal show higher levels of student achievement than other schools, after controlling for student characteristics and initial achievement levels. This involves comparing student achievement for students attending New Leaders-led OUSD schools with achievement for students at other OUSD schools. The within-district comparison is important because these schools share all the observable and unobservable influences associated with being in the same district, and share common teacher populations and recruiting pools and methods, which leads to lower likelihood of variation based on differences in policies. That said, OUSD worked closely with New Leaders during this time period, which could lead to spillover effects impacting all OUSD schools and potentially cause this analysis to understate the full impacts of New Leaders on OUSD as a whole.

Table III.1 shows the ELA and math results from our primary model, with test scores for individual students as the outcome measure, controlling for prior year test scores in both subjects. These results use the restricted definition of the New Leaders intervention group, restricting it to only students who were in their second or greater year of exposure to New Leaders, and schools that were in their third year or greater of New Leaders tenure. The first two columns show the coefficients and p-values from this analysis for the within-district comparison, for ELA outcomes, where New Leaders-led schools are compared to other schools in OUSD. The first coefficient is the treatment effect, with a value of 0.06, and a p-value of 0.078, indicating the results is significant at the 10 percent level. This indicates that OUSD students at New Leaders-led schools had ELA scores 0.06 standard deviations higher than other OUSD students, after controlling for prior student test scores and other student and school characteristics. Other coefficients include those on both prior subject test scores, as well as other student demographic characteristics.

The next two columns of Table III.1 show the within-district math results for the same model and comparison group, still using the restricted definition of the intervention group. Here the effect is larger, with a coefficient of 0.16 on the treatment variable. The p-value of 0.001 indicates this effect is statistically significant at the 1 percent level, indicating a high degree of confidence that the difference is greater than zero. This means that OUSD students at New Leaders-led schools had math scores that were 0.16 standard deviations higher than other OUSD students, after controlling for prior student test scores and other student and school characteristics.

Table III.1. Student-level regression results, restricted definition of intervention group

	ELA		Math	
	Coefficient	***p*-Value**	**Coefficient**	***p*-Value**
Intervention group (students: 2+ years, schools, 3+ years)	0.06	0.078	0.16	0.001
Prior year ELA score	0.66	0.000	0.19	0.000
Prior year math score	0.15	0.000	0.50	0.000
Black or African American	-0.53	0.000	-0.43	0.000
Hispanic	-0.50	0.000	-0.24	0.000
Asian	-0.37	0.000	0.14	0.067
Other	-0.39	0.000	-0.17	0.000
Gifted	0.32	0.000	0.36	0.000
Special education	-0.24	0.000	-0.14	0.000
Regression statistics				
Number of observations (student-years)	83,374		83,026	
Number of students	48,905		48,776	
R-squared	0.696		0.583	

Source: OUSD student-level data; New Leaders school-level data

Note: Results from a linear regression with standard errors clustered at the school level are reported. Results exclude student test scores that are greater than three standard deviations away from the mean. The intervention group students are those who attended a New Leaders-led school for at least two years, where a New Leaders-led school is defined as one that had a New Leader for at least three years. The comparison groups contain students who did not attend a New Leaders-led school. Student California Standardized Test (CST) scores are standardized using California means and standard deviations for each grade and year.

ELA = English language arts; OUSD = Oakland Unified School District

    Table III.2 shows the results from the same comparisons, but using an unrestricted definition of the New Leaders intervention group, which includes all students exposed to New Leaders, even if only for one year or at a school in its first year led by a New Leader. The results from this comparison are very similar to the previous results, yielding positive and statistically significant New Leaders effects, this time with magnitudes of 0.07 in ELA (significant at the 5 percent level) and 0.12 in math (significant at the 1 percent level).

Table III.2. Student-level regression results, unrestricted definition of intervention group

	ELA		Math	
	Coefficient	*p*-Value	Coefficient	*p*-Value
Intervention group (ever had a New Leader)	0.07	0.044	0.12	0.011
Prior year ELA score	0.66	0.000	0.19	0.000
Prior year math score	0.14	0.000	0.50	0.000
Black or African American	-0.53	0.000	-0.43	0.000
Hispanic	-0.50	0.000	-0.24	0.000
Asian	-0.37	0.000	0.14	0.064
Other	-0.40	0.000	-0.17	0.000
Gifted	0.32	0.000	0.36	0.000
Special education	-0.24	0.000	-0.14	0.000
Regression statistics				
Number of observations (student-years)	83,374		83,026	
Number of students	48,905		48,776	
R-squared	0.696		0.583	

Source: OUSD student-level data; New Leaders school-level data

Note: Results from a linear regression with standard errors clustered at the school level are reported. Results exclude student test scores that are greater than three standard deviations away from the mean. The intervention group contains students who ever had a New Leader. The comparison groups contain students who did not attend a New Leaders-led school. Student California Standardized Test (CST) scores are standardized using California means and standard deviations for each grade and year.

ELA = English language arts; OUSD = Oakland Unified School District

To interpret the size of these effects, consider that this measures the impact of a New Leaders-led school on student test-score growth in a single year, so if a student attends the school for multiple years, this effect would accumulate. We can construct an estimate of the cumulative effect of attending a New Leaders-led school for three years. Research suggests there is some fade-out of impacts over time, so we follow Chetty and colleagues (2011) in assuming that two-thirds of the impacts persist over time. So a student who attended a New Leaders-led middle school from grades 6 to 8 would experience a 0.12 cumulative growth advantage in ELA, and 0.32 cumulative growth advantage in math over the three-year period. For example, such an effect would move a student who was initially in the 40th percentile upon entering grade 6 to the 45th percentile in ELA at the end of grade 8, and from the 40th to the 50th percentile in math.

Another way of interpreting these effects is to compare New Leaders effect sizes to national norms regarding the amount of student academic growth that takes place in a given year (Bloom et al. 2008). Expressed this way, our impacts suggest that, on average, New Leaders schools produce approximately four months of extra learning growth in math after three years, and one and one-half months of extra learning growth in ELA after three years.

Another way to provide context for understanding the magnitude of these results is to compare these effects to the size of existing achievement gaps, either by race/ethnicity or by socio-economic status. In OUSD there is a gap of 0.73 standard deviations between the average math test scores of black students and white students; students eligible for reduced-price school meals have math scores that are an average of 0.79 standard deviations lower than other students. Comparing the size of those gaps to the estimated impacts of New Leaders, the size of the math effect produced by New Leaders schools after three years is equivalent to about 44 percent of the local black-white test score gap and 40 percent of the local achievement gap between higher- and lower-income students.

Another way to evaluate the size of these findings is to compare them with other students of impacts of principal training programs on student achievement. Little research exists in this area, but the few studies that do examine this question find effects of comparable magnitude or less. Corcoran and colleagues (2012), looking at effects of the Aspiring Principals Program in New York City (unaffiliated with New Leaders), found relatively small effects on student achievement growth, with coefficients in the 0.03 to 0.05 range. Martorell and colleagues (2010) examine impacts from the New Leaders program in a variety of locations, including OUSD. They find positive effects for the within-OUSD comparison of similar magnitude to those found here.

B. School-level analyses

We now turn to an analysis of school-level data. These data have a couple of important advantages over the student-level data used in the prior analysis. One advantage is that the school-level data go back to the 2002–03 school year, so we can examine changes in test scores over a longer period. The other advantage is that the school-level data are available for all schools in the state. Because the student-level data are unavailable for most districts, we were unable to make a comparison with a representative external sample matched for comparability with OUSD schools, a comparison we can make with the school-level data. While school-level comparisons cannot be used to examine growth for individual students or control for prior-year performance, we can look at changes in average test scores over longer periods of time by school, a descriptive analysis that can help inform the findings from the student-level analysis.

To create the best possible comparison group, we used propensity score matching to select a sample of CA schools that closely match OUSD schools in terms of baseline achievement levels and demographics, as described in the Methods Section. We then examine changes in average test scores over time for New Leaders-led schools, other OUSD schools, schools in the matched comparison sample, and finally the full sample of CA schools.

Table III.3 shows the average characteristics for each of these four samples of schools. The first two rows compare average ELA and match scores for each group in 2003. Each of the first three groups of schools have much lower average test scores than the state as a whole, with OUSD having the lowest average math scores. All three groups also have considerably more economically disadvantaged and minority students than the state as a whole.

In terms of evaluating which group of schools is the best fit as a comparison for the New Leaders-led schools, both the groups of other OUSD schools and the matched comparison sample are close fits in terms of initial achievement levels, particularly for ELA tests. New

Leaders-led schools have only 33 percent Hispanic, which is very similar to that for OUSD and the matched comparison sample.

One important question is whether a comparison of changes in school-level average test scores over time will show similar results to the findings from the student-level comparison of individual test score growth. Table III.4 examines this, showing changes in school average test scores in math and reading for each set of schools during three different periods, first from 2003 to 2007, then from 2007 to 2013, and finally from 2003 to 2013.

The first column shows how average test scores changed for New Leaders-led schools during these three periods. Looking at the middle set of rows, which show the changes from 2007 to 2013, the period covered by the student-level analysis, we can see that average ELA scores rose by 0.11 and average math scores rose by 0.17. For other OUSD schools, in the second column, average ELA scores fell by 0.17 during this period, and average math scores fell by 0.21. The differences between test score changes for New Leaders-led schools and those for other OUSD schools are statistically significant at the 10 percent level for ELA and the 5 percent level for math. Based on this comparison, it appears that the changes in school-average test scores are qualitatively similar to the results from the analysis of student-level growth.

The matched comparison sample of schools, represented in the third column, shows very little change in average test scores over the 2007 to 2013 period, with scores falling by 0.01 for ELA and rising by 0.05 for math. This suggests that a comparison of test score growth for New Leaders-led schools with growth for a matched comparison sample of schools would likely yield a positive effect for New Leaders-led schools. However, in this case the differences are not statistically significant.

Because the school-level data go back to 2002–2003, we can examine how average test scores changed over this longer period. The first set of rows of Table III.4 shows this comparison, with changes in average test scores from 2003 to 2007 for each of the four groups. Both the New Leaders sample and other OUSD schools show gains in average scores for ELA from 2003–2007, whereas the matched comparison sample shows basically no change in average test scores in ELA from 2003 to 2007. For math, the average scores drop slightly for the New Leaders sample. The last set of rows, comparing changes in average test scores over the entire period from 2003 to 2013, shows that the New Leaders schools have considerably higher changes than the other samples. This suggests that if student-level data went back further to include schools years back to 2002-2003, a comparison with outside districts would suggest a more positive effect of having a New Leaders principal on student achievement growth.

Table III.3. Comparison of key characteristics of matched and other comparison groups

Characteristic	New Leaders-led schools (3+ years)		OUSD			Matched comparison sample			Statewide		
	Mean	SD	Mean	SD	p-Value	Mean	SD	p-Value	Mean	SD	p-Value
Average ELA test score, 2003	-0.49	1.04	-0.62	1.00	0.658	-0.54	0.80	0.834	0.05	0.94	0.017
Average math test score, 2003	-0.20	0.88	-0.51	0.80	0.176	-0.39	0.74	0.370	0.04	0.97	0.305
Title I eligibility	0.88	0.33	0.86	0.35	0.825	0.90	0.30	0.843	0.69	0.46	0.081
Student-teacher ratio	18.08	2.27	20.67	17.96	0.556	23.43	38.84	0.573	20.33	5.94	0.119
Percentage white	0.09	0.16	0.08	0.18	0.852	0.09	0.13	0.913	0.38	0.29	0.000
Percentage black	0.37	0.21	0.48	0.29	0.137	0.43	0.28	0.425	0.08	0.12	0.000
Percentage Hispanic	0.33	0.26	0.32	0.27	0.962	0.34	0.24	0.822	0.42	0.29	0.198

Source: California Department of Education STAR Data, 2003; National Center for Education Statistics Common Core of Data, 2003; New Leaders data

Note: Elementary and middle schools are included. We matched OUSD schools to, at most, one comparison school within California, using data from 2003 on the following characteristics: school-level average ELA and math test scores, Title I eligibility, student-teacher ratio, and race/ethnicity distribution (percentage white, percentage black, and percentage Hispanic). Test scores are standardized using the means and standard deviations from the statewide sample for the relevant year by grade and subject. The reported p-values are from tests of the statistical significance of the difference in each measure between schools led by a New Leader for at least three years and each comparison group.

OUSD = Oakland Unified School District; ELA = English language arts

Table III.4. Changes in average ELA and math test scores in New Leaders-led schools (3+ years) and non-New Leaders schools

Characteristic	New Leaders-led schools (3+ years)		OUSD			Matched comparison sample			Statewide		
	Mean	SD	Mean	SD	*p*-Value	Mean	SD	*p*-Value	Mean	SD	*p*-Value
Change in average test scores, 2003–2007											
Change in average ELA test score	0.10	0.46	0.11	0.57	0.954	-0.02	0.39	0.280	0.01	0.34	0.252
Change in average math test score	-0.08	0.48	0.01	0.66	0.610	-0.14	0.45	0.662	0.02	0.65	0.515
N	17		53			71			6,378		
Change in average test scores, 2007–2013											
Change in average ELA Test Score	0.11	0.52	-0.17	0.48	0.058	-0.01	0.67	0.493	-0.04	0.43	0.156
Change in average math test score	0.17	0.47	-0.21	0.58	0.025	0.05	0.69	0.531	-0.03	0.54	0.133
N	16		38			63			6,065		
Change in average test scores, 2003–2013											
Change in average ELA test score	0.28	0.57	0.04	0.72	0.262	-0.02	0.61	0.083	-0.03	0.51	0.016
Change in average math test score	0.13	0.64	-0.06	0.75	0.364	-0.03	0.68	0.376	-0.01	0.75	0.460
N	16		38			63			6,049		

Source: California Department of Education STAR Data, 2003; National Center for Education Statistics Common Core of Data, 2003; New Leaders data

Note: Elementary and middle schools are included. We matched OUSD schools to at most one comparison school within California, using data from 2003 on the following characteristics: school-level average ELA and math test scores, Title I eligibility, student-teacher ratio, and race/ethnicity distribution (percentage White, percentage Black, and percentage Hispanic). Test scores are standardized using the means and standard deviations from the statewide sample for the relevant year by grade and subject. The reported *p*-values are from tests of the statistical significance of the difference in each measure between schools led by a New Leader for at least three years and each comparison group.

OUSD = Oakland Unified School District; ELA = English language arts

IV. CONCLUSIONS

The goal of this report is to provide additional evidence on the effect of attending a New Leaders-led school on student achievement, by comparing students attending those schools with other students, both within the same district and in other districts.

The results of our student-level analysis show greater growth in achievement in both math and ELA for students with New Leaders-trained principals, compared to other students in OUSD, with effect sizes of 0.06 in ELA and 0.16 in math. These results are statistically significant and equal to or greater than the effects found in other studies of principal training programs (Corcoran et al. 2012, Martorell et al. 2010), and are robust to a variety of different specifications and treatment group definitions. If the single-year growth improvement is aggregated to a three-year period, accounting for fade out in effects over time, it would correspond to a 0.12 cumulative effect for ELA and 0.42 cumulative effect for math, which would move a student who was in the 40th percentile upon entering grade 6 to the 45th percentile in ELA, and to the 50th percentile in math at the end of grade 8. This math effect is the equivalent of four months of extra learning growth for students in New Leaders-led schools.

Attempts to compare student-level growth for New Leaders students with a comparison group from outside OUSD were hampered by difficulty recruiting a closely matched comparison sample. The two external districts that did agree to participate, FUSD and SBCUSD, are significantly different from OUSD in both average initial achievement levels and average student demographics. Descriptive analysis of school-level changes in average test scores finds that from 2003 to 2013 New Leaders-led schools in OUSD had greater increases in average test scores in both ELA and math than either other OUSD schools or schools in the matched statewide comparison sample. These results are qualitatively consistent with the student-level analysis, suggesting that a well-matched sample of comparison schools could find positive effects from a student-level comparison, particularly if student-level data are available going back to before the 2006–2007 school year.

Few studies exist that have examined the effects of principal training on student achievement, particularly employing longitudinal student-level data to examine growth for individual students over time, so this study provides an useful contribution to the literature. Extending this study with more complete and well-matched student-level data would be a useful avenue for future research, as would examining the effects of principal training programs in other locations.

APPENDIX

Two school districts, in addition to OUSD agreed to provide us with student-level data: Fontana Unified School District (FUSD) and San Bernardino City Unified School District (SBCUSD). These two districts have a much higher percentage of Hispanic students than OUSD, with OUSD having a higher percentage of black students. Additionally, average initial achievement levels are quite different between OUSD and the other two districts, particularly in ELA, which makes these districts less than ideal as a comparison sample (See Table A.1).

The sample sizes for the comparison districts: the FUSD sample contained approximately 50,000 unique students (100,000 student-year observations with data on current and prior-year test scores), and the SBCUSD sample contained approximately 70,000 unique students (130,000 student-year observations with data on current and prior-year test scores).

We did an analysis comparing student achievement growth at New Leaders-led schools in OUSD with that for students in the two participating comparison districts, FUSD and SBCUSD. Although the differences in average student initial achievement levels and demographic characteristics make this a flawed comparison group, it still gives some indication of how student growth for New Leaders-led schools compares to a non-OUSD group.

Table A.2 show the results for ELA and math using the restricted definition of the intervention group, where the comparison group includes only students from the two external comparison groups, FUSD and SBCUSD. These results only compare students at New Leaders-led schools in OUSD with students from the two external districts, omitting other OUSD students. For this comparison group the coefficients on the treatment effects are small, 0.03 and -0.04 for ELA and math respectively, and both are statistically insignificant at the 10 percent level, indicating that student achievement growth for students in New Leaders-led schools during this period is not noticeably different from growth for students in FUSD and SBCUSD.

Table A.3 shows results using the unrestricted definition of the treatment group. Comparisons with this broader comparison group are also statistically insignificant at the 10 percent level.

Table A.1. Comparison of characteristics of comparison groups

Characteristic	New Leaders-led schools (3+ years)		OUSD			FUSD and SBCUSD		
	Mean	SD	Mean	SD	p-Value	Mean	SD	p-Value
Average ELA test score, 2003	-0.49	1.04	-0.62	1.00	0.658	-0.82	0.46	0.044
Average math test score, 2003	-0.20	0.88	-0.51	0.80	0.176	-0.47	0.59	0.132
Title I eligibility	0.88	0.33	0.86	0.35	0.825	0.98	0.16	0.080
student-teacher ratio	18.08	2.27	20.67	17.96	0.556	21.53	1.67	0.000
Percentage white	0.09	0.16	0.08	0.18	0.852	0.14	0.09	0.053
Percentage black	0.37	0.21	0.48	0.29	0.137	0.16	0.09	0.000
Percentage Hispanic	0.33	0.26	0.32	0.27	0.962	0.67	0.15	0.000

Source: California Department of Education STAR Data, 2003; National Center for Education Statistics Common Core of Data, 2003; New Leaders data.

Note: Elementary and middle schools are included. Test scores are standardized using the means and standard deviations from the statewide sample for the relevant year. The reported p-values are from tests of the statistical significance of the difference in each measure between schools led by a New Leader for at least three years and each comparison group.

OUSD = Oakland Unified School District; FUSD = Fontana Unified School District; SBCUSD = San Bernardino Unified School District; ELA = English language arts

Table A.2. Student-level regression results from comparison with FUSD and SBCUSD, restricted definition of intervention group

	ELA		Math	
	Coefficient	***p*-Value**	**Coefficient**	***p*-Value**
Intervention group (students: 2+ years, schools, 3+ years)	-0.03	0.453	-0.04	0.541
Prior year ELA score	0.69	0.000	0.20	0.000
Prior year math score	0.15	0.000	0.54	0.000
Black or African American	-0.13	0.000	-0.11	0.000
Hispanic	-0.11	0.000	0.04	0.106
Asian	0.01	0.675	0.36	0.000
Other	-0.04	0.078	0.07	0.009
Gifted	0.24	0.000	0.34	0.000
Special education	-0.27	0.000	-0.08	0.000
Regression statistics				
Number of observations (student-years)	240,490		240,075	
Number of students	126,424		126,562	
R-squared	0.682		0.550	

Source: OUSD, FUSD, and SBCUSD student-level data; New Leaders school-level data

Note: Results from a linear regression with standard errors clustered at the school level are reported. Results exclude student test scores that are greater than three standard deviations away from the mean. The intervention group contains students who had a New Leader for at least two years who attended a school that was led by a New Leader for at least three years. The comparison groups contain students who did not attend a New Leaders-led school. Student California Standardized Test (CST) scores are standardized using California means and standard deviations for each grade and year.

ELA = English language arts; OUSD = Oakland Unified School District; FUSD = Fontana Unified School District; SBCUSD = San Bernardino City Unified School District

Table A.3. Student-level regression results from comparison with FUSD and SBCUSD, unrestricted definition of intervention group

	ELA		Math	
	Coefficient	*p*-Value	Coefficient	*p*-Value
Intervention group (ever had a new leader)	-0.04	0.288	-0.07	0.174
Prior year ELA score	0.69	0.000	0.20	0.000
Prior year math score	0.15	0.000	0.54	0.000
Black or African American	-0.14	0.000	-0.13	0.000
Hispanic	-0.12	0.000	0.03	0.192
Asian	0.01	0.781	0.38	0.000
Other	-0.04	0.030	0.06	0.024
Gifted	0.25	0.000	0.35	0.000
Special education	-0.27	0.000	-0.08	0.000
Regression statistics				
Number of observations (student-years)	250,711		250,230	
Number of students	135,600		135,648	
R-squared	0.682		0.552	

Source: OUSD, FUSD, and SBCUSD student-level data; New Leaders school-level data

Note: Results from a linear regression with standard errors clustered at the school level are reported. Results exclude student test scores that are greater than three standard deviations away from the mean. The intervention group contains students who ever had a New Leader. The comparison groups contain students who did not attend a New Leaders-led school. Student California Standardized Test (CST) scores are standardized using California means and standard deviations for each grade and year.

ELA = English language arts; OUSD = Oakland Unified School District; FUSD = Fontana Unified School District; SBCUSD = San Bernardino City Unified School District

REFERENCES

Bloom, Howard, Carolyn Hill, Alison Rebeck Black, and Mark Lipsey. "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions." New York: MRDC, October 2008.

Chetty, Raj, John Friedman, and Jonah Rockoff. "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." NBER Working Paper. Cambridge: NBER, December 2011.

Corcoran, Sean P., Amy Ellen Schwartz, and Meryle Weinstein. "Training Your Own: The Impact of New York City's Aspiring Principals Program on Achievement." *Educational Evaluation and Policy Analysis*, vol. 34, no. 2, 2012, pp. 232-253.

Gates, Susan, Laura Hamilton, Paco Martorell, Susan Burkhauser, Paul Heaton, Ashley Pierson, Matthew Baird, Mirka Vuollo, Jennifer Li, Diana Lavery, Melody Harvey, and Kun Gu. "Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in 11 Districts." Santa Monica, CA: RAND, March 2014.

Martorell, Paco, Paul Heaton, Susan Gates, and Laura Hamilton. "Preliminary Findings from the New Leaders for New Schools Evaluation." Santa Monica, CA: RAND, 2010.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC