



EDUCATION

CHILDREN AND FAMILIES
EDUCATION AND THE ARTS
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INFRASTRUCTURE AND
TRANSPORTATION
INTERNATIONAL AFFAIRS
LAW AND BUSINESS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
TERRORISM AND
HOMELAND SECURITY

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from www.rand.org as a public service of the RAND Corporation.

Skip all front matter: [Jump to Page 1](#) ▼

Support RAND

[Browse Reports & Bookstore](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Education](#)

View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND electronic documents to a non-RAND website is prohibited. RAND electronic documents are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).

This product is part of the RAND Corporation occasional paper series. RAND occasional papers may include an informed perspective on a timely policy issue, a discussion of new research methodologies, essays, a paper presented at a conference, a conference summary, or a summary of work in progress. All RAND occasional papers undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

OCCASIONAL PAPER

Addressing Challenges in Evaluating School Principal Improvement Efforts

Susan Burkhauser, Ashley Pierson, Susan M. Gates, Laura S. Hamilton

Sponsored by New Leaders

This work was sponsored by New Leaders. The research was conducted in RAND Education, a unit of the RAND Corporation.

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2012 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND website is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page (<http://www.rand.org/publications/permissions.html>).

Published 2012 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: <http://www.rand.org>
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This report highlights challenges that states, districts, and other entities can expect to encounter as they evaluate efforts to improve school leadership and presents recommendations to mitigate these challenges. The study draws on lessons learned during the RAND Corporation's multiyear evaluation of the New Leaders program. Since 2006, New Leaders has contracted with the RAND Corporation to conduct a formative and summative evaluation of the program, its theory of action, and its implementation. New Leaders is a nonprofit organization dedicated to promoting student achievement by developing school leaders to serve in urban schools.

The recommendations described here will be of interest to policymakers in school districts, charter management organizations (CMOs), state education agencies, evaluators of efforts to improve school leadership, and data management personnel.

This research was conducted in RAND Education, a unit of the RAND Corporation, under a contract with New Leaders. Additional information about RAND Education can be found at <http://www.rand.org/education>.

Contents

Preface	iii
Acknowledgments	vii
Abbreviations	ix
CHAPTER ONE	
Introduction	1
CHAPTER TWO	
RAND's Evaluation of the New Leaders Program	5
Program Overview	5
Student Outcome Analysis	5
Additional Components of the Evaluation	6
CHAPTER THREE	
Challenges in Using Outcome Data to Evaluate School Leadership Improvement Efforts	7
Using Student Outcome Measures	7
Inconsistency in Outcome Measures	7
Measure Manipulation	9
Tracking Students Across Districts	9
Lack of Adequate High School Outcome Measures	10
Effects of Student Dropout	11
Timing of Data and Impact	12
Controlling for Student Characteristics	14
Unobserved Characteristics	14
Observed Characteristics	15
Accounting for School Context	16
Determining Appropriate Comparison Schools	16
Developing Measures of School Context	18
Measuring Principal Impact in Differing Contexts	19
Controlling for Principal Characteristics	20
Quality and Availability of Principal Tenure Data	20
Variation in Principal Career and Training Paths	21
Incorporating Multiple Measures	22
CHAPTER FOUR	
Conclusion	25
References	27

Acknowledgments

This paper summarizes key insights that RAND Education has developed about the evaluation of efforts targeting principals through our multiyear effort to evaluate the New Leaders program. These insights have emerged through the work of a large team, not all of whom are reflected in the list of authors for this paper. In particular, Paco Martorell, who leads the analysis of student achievement data for this evaluation, is the source of many of the key points raised here. We also acknowledge the contributions of Paul Heaton and Mirka Vuollo, key members of the evaluation team. New Leaders staff Gina Ikemoto, Brenda Neuman-Sheldon, Ben Fenton, Lori Taliafero, and Jackie Gran provided useful feedback on the overall development of this paper, as well as on earlier drafts. We are also grateful to Cathy Stasz of RAND, who provided helpful comments on an early draft, and to Kerri Briggs at the George W. Bush Institute and RAND colleague John Engberg, who reviewed the report and provided constructive suggestions for improvement.

Donna White helped to compile and format the final document. Nora Spiering edited the final copy. The authors take full responsibility for any errors.

Abbreviations

ACS	American Community Survey
AEFP	Association for Education Finance and Policy
AP	Advanced Placement
APP	Aspiring Principals Program
AYP	Adequate Yearly Progress
CMO	charter management organization
FARM	free and reduced-price meal
FR	Federal Register
GED	General Educational Development
GPA	grade point average
IEP	individualized education program
NCLB	No Child Left Behind
SES	socioeconomic status

Introduction

Effective school leaders are a critical component of what makes a school successful. The role that school principals and other leaders play in improving the performance of schools is a burgeoning field of research. State and district policymakers, as well as other organizations, such as foundations and nonprofits, are emphasizing efforts targeting school leadership as a way to improve student outcomes. Given the focus on accountability in education, policymakers and funders are keenly interested in evaluating whether efforts aimed at improving school leadership result in improved student learning.

The efforts designed to improve school leadership include a wide range of policies, practices, and programs undertaken by states, districts, and charter management organizations (CMOs), as well as organizations that do not provide direct services to students (e.g., independent principal preparation programs or foundations). Principals, who have primary responsibility for leading schools, are the target of many of these efforts. These include efforts to improve the skills and competencies of current and future principals, the way schools and districts manage principals, and the environments in which principals work. The efforts may involve new activities or reforms to current policies and could be implemented at the state or district level. Potential efforts are the provision of coaching for new principals; greater autonomy for principals; the training of aspiring principals; and new approaches to the selection, placement, and provision of professional development for new or current principals. These efforts might span multiple states or districts or be implemented by CMOs or other organizations with an interest in principal leadership. Often such efforts are introduced without incorporating formal methods for their evaluation, in spite of the fact that it is important to understand whether the efforts work and are a good use of resources.

In the current era of accountability, gains in student achievement are the key criteria against which stakeholders seek to judge the effect of these efforts. The evaluation of these school leadership improvement efforts is distinct from evaluating individual principal performance, although the measures used for individual principal performance evaluation could also be used for the broader evaluation of improvement efforts. The federal No Child Left Behind (NCLB) Act of 2001 required all public schools to administer standardized tests and to issue public reports of school-level test scores each year. Failure to meet the test-score targets set by states leads to an escalating set of sanctions and interventions. As a result of this law, district and school administrators have increased their emphasis on raising student achievement. Recently, the federal government allowed states to apply for waivers to provide flexibility for the 2014 proficiency target. One requirement for receiving a waiver is that the state must submit plans to develop systems for evaluating and supporting teacher and principal effective-

ness that include multiple performance measures, including measures of student progress (U.S. Department of Education, 2011).

The use of multiple performance measures is becoming standard practice in evaluation for both teachers and school leaders. Recently, many school districts and states have included multiple student achievement measures as a component of their principal performance evaluation methods.¹ Additionally, there is a growing literature on the use of student achievement measures to evaluate school leadership improvement efforts. As the pathways from improved school leadership to improved student outcomes are both indirect and diverse, the development of metrics for assessing the success of efforts to improve school leadership poses important challenges.

Over the past five years, RAND Education, a unit of the RAND Corporation, has been engaged in an evaluation of the New Leaders Aspiring Principals program (hereafter referred to as New Leaders). New Leaders is a nonprofit organization dedicated to promoting student achievement by developing school leaders to serve in urban schools. Through this project, the researchers have gained practical experience in the issues involved in evaluating efforts that are designed to improve school leadership. The lessons highlighted here are derived from that experience.

In this report, we describe the challenges that states, districts, and other entities can expect to encounter as they evaluate efforts to improve school leadership and offer suggestions for dealing with those challenges based on our project experience and understanding of the literature. We focus on lessons learned pertaining to the use of student achievement and other administrative data for the purpose of evaluating efforts that seek to improve the leadership of school principals. We do not address all of the challenges associated with evaluating school leadership but instead focus on topics that are relevant to the use of student outcomes and school and principal data in those evaluations. The discussion in this report applies to evaluations of policies, practices, and programs and not to individual principal performance evaluations, although some of the issues that arise in the evaluation of individual principals may pertain to the evaluation of efforts to improve school leadership as well.

This report is intended for district and state policymakers who are addressing school leadership issues and for others who are tasked with evaluating an effort to improve school leadership. Evaluators in this context could be employees of a school district, CMO, or state or could be part of an external organization, such as a funding or implementing agency. These efforts include pre-service principal preparation programs, as well as programs that provide ongoing coaching or other forms of support. Employees of these programs are viewed as evaluators in this context—whether they are using the results of their evaluation for program improvement purposes, for monitoring of outcomes, or for reporting to funders. Not all of the recommendations we present in this report will be directly relevant to all efforts being carried out, nor will they address the challenges that arise in all situations.

First, we discuss challenges involved in using student outcome measures as part of an evaluation. Next, we weigh concerns and cautions that arise when evaluations need to control for student characteristics. Then, we describe how the school context functions as an important mediator between leadership efforts and effects on student outcomes and discuss the challenges

¹ From the beginning of 2010 to June 2011, Arizona, California, Connecticut, Idaho, Illinois, Indiana, Maine, Maryland, Michigan, Nevada, and Ohio adopted legislative changes to include student achievement as an optional or a required portion of principal evaluations (Piro, Wiemers, and Shutt, 2011).

involved in appropriately accounting for that context. We then examine principal characteristics as a potential confounding factor in evaluations of leadership efforts and the challenges involved in accounting for principal characteristics. Finally, we discuss the importance of using multiple measures to evaluate efforts to improve school leadership.

RAND's Evaluation of the New Leaders Program

Program Overview

New Leaders is dedicated to promoting student achievement by developing outstanding school leaders to serve in urban schools. In support of this objective, New Leaders developed a model, or “theory of action,” of the relationship between effective school leadership and improved student achievement. It then designed and implemented a program based on that model to recruit, train, and support school leaders. The New Leaders organization has partnered with a number of major urban school districts and CMOs to recruit, select, and train principals to serve in high-needs schools. These partners are located in nine different states and in Washington, D.C. New Leaders principals have been placed in a wide range of schools throughout the partner districts, with both traditional and atypical grade-level configurations. New Leaders principals have been placed in charter schools, traditional district schools, start-up schools, turnaround schools, and schools with a special focus.¹ RAND Education is conducting a multiyear formative and summative evaluation of the New Leaders program, its theory of action, and its implementation. This evaluation is sponsored by New Leaders.

Student Outcome Analysis

Our evaluation incorporates an annual student outcome analysis that uses student-level data in tandem with information about principals and schools to produce estimates of the program's effect on student outcomes (Martorell et al., 2010). The most recent analysis (the fifth conducted to date), completed in August 2011, incorporates data from seven school districts through school year 2009–2010². RAND plans to conduct additional analyses in 2012 and 2013 of data from school years 2010–2011 and 2011–2012, respectively.

To estimate the program's effect on student outcomes, we used several modeling approaches and controlled for various student-level and school-level characteristics that may affect outcomes.³ We examined the program's effect on standardized test scores in mathematics and reading in all districts and on a variety of other student outcomes, including attendance,

¹ For more information on the New Leaders program, please see their website at <http://www.newleaders.org/>.

² The 2010–2011 school year data were not available at the time of analysis, as there is often a delay of a year or more in receiving test score data from districts and CMOs. Additionally, the span of data available varies by district; some districts provide historical data going back to school year 2001–2002. Others provide fewer school years of data, depending on the year the New Leaders program began in that district.

³ We estimated fixed effects, random effects, and first-difference models.

dropout, and graduation, depending on the availability of data in the district.⁴ Different states use different achievement tests, generating scores that are not directly comparable. In order to combine results from different states in a single analysis, we normalized the standardized test scores.⁵ We used these normalized data to generate national program effect estimates. In addition, we estimated program effects by district. In both the district and national estimations, we performed separate analyses for lower grades (K–8) and upper grades (9–12). Testing occurs more frequently during lower grades, and thus we were able to use multiple years of data in our models; for upper grades, we typically estimated cross-sectional models using a single year of data.

We used administrative data on students, schools, and principals provided by the school districts to perform this analysis.⁶ The main advantage to using administrative data rather than publicly available data is that it provides information on individual students, which allowed us to control for student factors that may affect student outcomes, such as free and reduced-price meal (FARM) status. The inclusion of student-level data for multiple years also permits controlling for outcomes in prior years and the removal of student factors that are constant over time. This student-level data over multiple years helps to estimate the unique effect of the program and improves the accuracy of the evaluation.

Additional Components of the Evaluation

In addition to the annual student achievement analyses, we administered and analyzed principal surveys in the 2007–2008 and 2010–2011 school years. In the 2008–2009 and 2009–2010 school years, we carried out case studies of principals who were in their first year of principalship in 2008–2009, following them into their second year.⁷ Our analysis of both qualitative and administrative data has provided us with useful insights into the strengths and weaknesses of different data sources, as well as the limitations of available data. For example, when analyzing data from the principal surveys and comparing these data with administrative data, we discovered inconsistencies in principal tenure at the school level, which is an important variable used to examine the impact of principals on student outcomes. Additionally, the case studies provided evidence that the school environment in which a principal is placed varies greatly and can affect the influence that a principal has on his or her school. Another aspect of our evaluation of New Leaders is the annual yearly interviews with key district administrators. These interviews provide context that help us interpret the results of the student achievement and survey data analysis.

⁴ Certain non-test student outcomes are not available in all districts.

⁵ That is, we converted test scores to a scale with a common mean and standard deviation.

⁶ In some cases, we receive data from the state, CMO, or governing board.

⁷ We completed a study on principals in their first year at a given school using a combination of survey data, case study data, and administrative data (please see Burkhauser et al., 2012).

Challenges in Using Outcome Data to Evaluate School Leadership Improvement Efforts

This chapter discusses challenges associated with evaluating efforts to improve school leadership. The statistical models in the New Leaders evaluation enable us to estimate the overall effect of the New Leaders program on selected student outcomes, independent of the effect of a set of control variables. It is this overall, aggregate estimate that informs whether the program is having an effect on the student outcomes.

Using Student Outcome Measures

The available outcome measures for the evaluation of efforts to improve school leadership typically include students' scores on state or district assessments, along with other student-level information, such as whether a student graduated or progressed to the next grade. Student outcome data are critical to understanding how well an effort is working to improve the principalship, but they have a number of limitations. Below we discuss six broad issues that many evaluations are likely to encounter.

Inconsistency in Outcome Measures

Challenge: When evaluating efforts to improve school leadership, there are often inconsistencies in the availability of outcome measures across states, districts, and CMOs. Even within a district, changing reporting needs and inadequate record-keeping can lead to differences in the availability of outcome measures from year to year. Depending on whether charter school data are reported separately or by the district, charter schools may track different outcomes than the district with which they are affiliated. These inconsistencies make it challenging to accurately evaluate the improvement effort, as the same outcomes are needed to compare results between years and across districts, states, and CMOs. Such inconsistency may require the estimation of separate program effect measures for different districts, states, or types of schools. This limits the interpretation of results derived from each type of estimation, as the results may vary greatly and may not be generalizable beyond the specific district, state, or school type.

For example, for the New Leaders evaluation, some districts provided us with detailed attendance data that included days attended and days enrolled at each school that a student attended throughout a given school year. These detailed data allowed us to create an average attendance rate using information from all schools that a student attended. Other districts initially provided attendance data that only included information from the school at which the

student spent the most time, which is not equivalent to the average attendance rate for students who attend more than one school during the school year. In an attempt to resolve this inconsistency in attendance rates for students who switch schools, we requested detailed attendance data, where districts were able to provide it, for each student at each school they attended in order to make a comparable attendance measure across districts.

Recommendation for Policymakers: We recommend that districts (and states, if applicable) ensure that the same indicators are tracked year to year to the extent possible. When changes are made, they should be clearly documented. States should consider requiring certain important indicators in annual reports from both districts and CMOs; this would facilitate cross-district evaluations and evaluations that include charter schools. An agreement between districts and states on how to standardize certain variables would be useful prior to implementing any statewide or cross-district effort targeting school leadership.

Recommendations for Evaluators:

1. *Meta-analysis:* One solution is to combine the separate effect estimates from each state, district, or CMO using a meta-analysis procedure, which would provide an overall estimate of the effort's effect.
2. *Normalization and standardization:* When the evaluation encompasses multiple districts or CMOs in different states, we recommend using normalized test scores when combining all observations into one analysis.¹ Where feasible, other outcomes and measures should also be standardized across districts and CMOs.² For districts using the same test (e.g., those within the same state), normalization of test scores is not as essential in conducting the analysis. It may still be desirable, however, because normalized measures are often easier to interpret. Additionally, states may change their tests over time, so a multiyear analysis of achievement scores from one state may require normalization to be comparable across years. For analyses of non-test outcomes (e.g., dropout rate, attendance rate, and suspensions) in one state combining data from multiple districts, standardization of the definitions of these outcomes at the state level would facilitate analysis. Separate, district-specific analyses for various outcome measures can be used where standardization is not possible. For an evaluation of an effort taking place in a single district or CMO, standardization across schools within the district or CMO is important.
3. *Choice of outcome measures:* Lastly, evaluators should weigh carefully the decision of which outcome measures and results to emphasize over others. Student achievement metrics are most frequently emphasized, but additional outcome measures, such as grade promotion, attendance, graduation, and postsecondary outcomes should also be considered.³ In selecting and prioritizing metrics, evaluators will want to consider not only data reliability, validity, and availability, but also program objectives. This decision

¹ In this report we define *normalized* test scores as scores that have been put on the same scale (e.g., test scores have a mean of zero and a standard deviation of one).

² In this report we use the term *standardized* to refer to consistency in general (i.e., variables are defined in the same way across districts and CMOs).

³ Grade promotion is typically an indicator variable marking whether a student advanced to the next grade at the end of the school year. Retention is a related indicator marking whether the student did not advance.

will also depend partially on stakeholder interests, as some stakeholders will be more interested in test scores or in interim measures than other outcomes.

Measure Manipulation

Challenge: Measures being used to evaluate efforts to improve school leadership might be subject to manipulation, particularly if they are used for high-stakes purposes, such as the evaluation of individual principals or other administrators. Principals and other administrators have an incentive to focus on improving measures that contribute to their own evaluation. Research suggests that when high stakes are attached to performance measures, those being evaluated for performance are likely to shift time away from activities that are not captured by the high-stakes measures and to engage in practices that can inflate or otherwise distort the high-stakes measures (Koretz, 2008; Hamilton, Stecher, and Yuan, 2012). The possibility of these types of responses raises concerns about the validity and utility of such measures for evaluating an improvement effort. Additionally, if the district has a stake in the school leadership improvement effort, district officials will want to be able to show stakeholders that the effort is a success. This can result in district officials manipulating the measures used to evaluate the effort.

Recommendation for Evaluators: We recommend that evaluators try to determine what measures are being used in personnel evaluations of principals or other administrators when choosing measures to use in the evaluation of an effort to improve school leadership. At a minimum, the evaluation effort should investigate whether low-stakes outcomes (which might include retention or attendance, depending on the district) provide evidence that is consistent with the information provided by high-stakes outcomes (such as test scores or dropout rates). Evaluators should also consider the likelihood that the evaluation of the overall effort could be biased by measure manipulation by principals. It is also important to determine whether district officials can manipulate the measure. If possible, steps should be taken to audit or monitor opportunities for manipulation at both the school and district levels.

Tracking Students Across Districts

Challenge: The inability to track students across districts or between districts and CMOs is a problem because the evaluation should control for prior student outcomes to help isolate the effect of the effort. District-to-CMO tracking is only an issue in districts in which charter schools have separate record-keeping from the school district (in some states, charter schools are not run by the state). If a student transfers across districts, the data usually include a marker to indicate the transfer but often do not include accompanying information about prior achievement or other outcomes.

Some states have created a statewide database that provides each student with a unique identification number, allowing districts and CMOs in the state to follow transfer students.⁴ These data systems provide access to prior test scores even when students transfer.

Recommendation for Policymakers: We recommend that other states consider developing systems that enable tracking of students throughout the districts and CMOs operating in their jurisdiction.

⁴ For more information on state data quality, please see the Data Quality Campaign website at <http://www.dataqualitycampaign.org/>. This organization maintains and updates a list of features of each state's data system.

Recommendation for Evaluators: In the absence of transfer data, the evaluator may wish to control for the lack of prior test scores for this specific group of students. This can be done by creating a variable that marks transfer students as not having prior test scores.⁵

Lack of Adequate High School Outcome Measures

Challenge: Assessing principal performance at the secondary level is especially challenging because statewide achievement tests are administered less frequently and consistently than at the elementary and middle school levels. The NCLB Act of 2001 mandated annual testing of students in mathematics and reading for grades 3 through 8 and one annual test for students in grade 10, 11, or 12 in each of three subjects. The availability of annual data for all students in grades 3 through 8 allows evaluators to control for student learning in previous grades prior to the year being evaluated by incorporating the test scores that student received the previous year (for those in grades 4 through 8). This helps to isolate the effect of the principal in the year being evaluated.⁶

At the high school level, however, some states test students only once while in grades 10 through 12.⁷ In these states, controlling for the prior achievement of individual students is much more difficult. As a result, it is more difficult to isolate the effect of the principal at the school.

Recommendations for Evaluators:

1. *Course-taking:* If data on students' course-taking behavior are available, the number of credits earned and rigor of curriculum (i.e., honors or Advanced Placement [AP] courses) could be used as student outcome or control measures. These measures could be used in tandem with others and may be particularly useful in the absence of student test score data. Another measure that could be constructed from course-taking data is course repetition. At the school level, this could be measured as the percentage of students who have had to repeat classes. At the student level, this could be how many classes the student has had to repeat or an indicator variable for whether the student has repeated a class.
2. *Other tests:* Evaluators may be able to access results from other tests to assess student outcomes at the student level. These include required end-of-course tests, exit exams for high school graduation, and tests taken for college entry, such as the SAT or ACT. These alternative tests may be used as student outcome measures or as a control for student performance when assessing principals, but in many cases adjustments would need to be made to address the self-selected nature of the populations of students tested. Evaluators need to consider other limitations of these tests, such as a possible lack of alignment to the school curriculum in the case of the SAT, ACT, and exit exams, and whether the tests are appropriate measures of the principal's effect on student learning.

⁵ This could be an indicator variable, with a value of 1 for transfer students with no prior test scores and a value of 0 for those with prior test scores in the system.

⁶ This can be achieved through the use of value-added measures; please see Lipscomb et al. (2010) for a detailed discussion of this method.

⁷ Some states and districts test more than once during a student's high school career; for example, California administers mathematics, reading, and other subject tests each spring in grades 2 through 11.

3. *Additional high school outcome measures:* Additional student outcome measures that can be used to assess the performance of a principal include attendance, school transfer rates, and suspensions.⁸ Another alternative outcome measure specific to high schools is graduation rate. Rumberger and Palardy (2005) suggest that multiple alternative measures, such as dropout and school transfer rates, should be used with student test scores to judge performance at the high school level.
4. *Postsecondary outcomes:* Postsecondary outcomes, such as college entrance, persistence, and degree completion, can also be used to assess high school principal efforts.⁹ While the use of such outcomes can provide important insights into the long-run implications of efforts targeting principals, policymakers should consider the amount of time that must pass before such measures can be used. For example, with a principal preparation program, over a decade could pass between the time that an aspiring leader is trained, placed, and leads a school and the time that students in the school graduate and attend college.
5. *Multiple outcome measures:* The RAND evaluation of New Leaders uses such high school outcomes as retention, dropout, graduation, AP course enrollment, attendance, progression from 9th to 10th grade, and credits taken. We recommend that multiple outcome measures (such as attendance, dropout and graduation, transfer rate, and college enrollment) be used in addition to test scores. Because the available measures and the ability to create standardized measures will likely vary by district, evaluators of efforts that span multiple districts will need to weigh the trade-offs between using additional measures and including all districts in the analysis. The decision should be based on a number of factors, including how many principals are to be evaluated in each district and the availability of consistent measures across districts.

Effects of Student Dropout

Student dropout from high school becomes a serious issue for the evaluation of principals. While some students drop out of school during middle school (see, for example, Rumberger, 1995), dropout is much more common in high school, when students reach and surpass the compulsory schooling age. For example, in 2010, California's eighth grade dropout rate was reported as 3.49 percent, compared with 18.2 percent at the high school level (Bonsteel, 2011).

Preventing dropout and encouraging graduation is typically one of the goals of a secondary school administrator; some districts even include graduation rate in their high school principal performance assessments. For example, Chicago Public Schools includes school graduation rate as compared with the district average as a competency in their scoring rubric for principal evaluations. Charlotte-Mecklenburg Schools sets a graduation rate target for performance evaluation of principal supervisors.

⁸ These alternative measures can be used at other school levels, although they are most pertinent to high schools, given the issues with a lack of annual achievement test scores for each high school student discussed above. These measures may also be helpful in evaluating efforts that involve middle school students (Balfanz, Herzog, and Mac Iver, 2007).

⁹ For example, an evaluation of the Milwaukee school voucher program used enrollment in a four-year college and college persistence as outcome measures (Wolf, 2012). Booker et al. (2011), using data from the Florida Department of Education's K–20 Education Data Warehouse, Chicago Public Schools, and the National Student Clearinghouse, examines attendance at two- or four-year colleges or universities within five years of high school completion.

Challenge 1: Students who drop out of high school often perform worse at school than their peers who continue and graduate (Alexander, Entwisle, and Horsey, 1997; Rumberger and Thomas, 2000). If a principal encourages students to stay in school, the average standardized test scores in his or her school may decrease because of the presence of these low performers in the pool of assessed students. This makes a principal appear less effective when performance is evaluated using the common outcome measure of student test scores. However, the same principal would show gains in graduation rates and decreases in dropout rates.

Recommendation for Evaluators: Because the outcome measure(s) used for principal evaluations may present conflicting incentives, we recommend that multiple student outcome measures should be used at the high school level. To avoid penalizing principals who encourage students to stay in school, potentially lowering average school test scores, multiple measures that include graduation (or dropout) rates and test scores should be used to evaluate school leadership improvement efforts.

Recommendation for Policymakers: Student dropout and graduation should be tracked as much as possible. Dropout can prove challenging to monitor, as students who transfer to other schools or districts may appear in the data to have dropped out. For example, students may leave a traditional school to attend an alternative program, such as a General Educational Development (GED) program, which may not trigger the request for student records that a transfer to another school typically would. The request for student records serves as a prompt to districts to record in the database that the student transferred to another school. GED students should not be counted as dropouts but as transfers to an alternative program.

Challenge 2: Graduation rates can be difficult to calculate; various formulas have been used. Part of the complication in constructing a graduation rate measure is how to include students who complete high school in five years (or more) or GED recipients as graduates in the formula. For a discussion of graduation rate calculations, see Hauser and Koenig (2011).

Recommendation for Policymakers: We recommend that districts attempt to track and categorize student movement and dropout accurately to assist in the calculation of graduation rates, as well as proper attribution of student test scores to certain schools and principals. In 2008, the federal government released regulations (see 73 FR 64435–64513) to govern how graduation rates are calculated and reported for NCLB purposes. These regulations require reporting the four-year adjusted cohort graduation rate beginning in the 2010–2011 school year and also require school officials to have written confirmation of students leaving the cohort because of transfers or other causes. These regulations should be followed to ensure that graduation rates are calculated using the same method across districts, CMOs, and states.

Recommendation for Evaluators: States may request additional time to report the four-year adjusted cohort graduation rate. Also, it might be unclear from the data which calculation is used for a district- or state-provided graduation rate. Evaluators should calculate their own graduation rate (using an appropriate definition) using student-level dropout and graduation data, if possible, to ensure that the district- or state-provided rate is calculated correctly and consistently.

Timing of Data and Impact

Principals influence student outcomes mainly through their influence over teachers. Teachers, in turn, directly affect student outcomes through their classroom contact with students.

Challenge 1: This indirect relationship between principals and students may cause a delay between the start of the school leadership intervention and any sizable changes in student outcomes.

Recommendation for Policymakers and Evaluators: It is crucial for evaluators and policymakers to recognize that the time frame for observing a measurable effect on student outcomes from efforts to improve principals may be substantially longer than for efforts that target teachers or students directly. This is particularly true of programs focused on leadership training because the aspiring leaders must go through the training, be placed as principals, and serve as principals for some length of time before student outcomes should be examined.

Recommendations for Evaluators:

1. *Multiple years:* It is preferable that the evaluation include at least two years of data from schools in which principals are placed. Even then, it is important to recognize that two years might not be enough time to detect a principal's effect on student outcomes.
2. *Interim measures:* Interim measures could be used throughout the course of the evaluation to determine if the improvement effort is "on track" to affect student outcomes. Principals affect many aspects of their schools, some of which can be assessed and used to track progress on improvement efforts. Examples of interim measures are changes in school culture, including staff sharing the same vision and goals for the school, emphasis on college readiness for every child, establishment of a uniform student disciplinary system, creating an anti-bullying program, and ensuring order and safety in the school building. Changes in teacher practices, including data-driven instruction, collaborating with colleagues, or feeling a sense of urgency to improve student outcomes, could also be interim measures. Principal practices that could be used as interim measures include time use; communication with teachers, parents, and other members of the school community; and creating clear and consistent rules and professional expectations for staff. The evaluator should articulate the theory of action that delineates what is expected to happen as a result of the school leadership improvement effort. The evaluator then should determine which interim measures would be expected to show changes and at what point in time these changes would be expected to take place. For example, a principal improvement program aimed at increasing student achievement scores may wish to use changes in the composition of the teaching staff over time (such as retention of high-value-added or highly rated teachers) and records of principal time spent with teachers (particularly with low-scoring or novice teachers) as interim measures, based on the expectation that changes in student achievement would result from the principal's management of the school's human capital stock. However, the evaluator may want to consider whether such an expectation is appropriate for all types of schools. For example, in a larger school or a high school, principals might delegate teacher coaching and observation activities. In this case, an alternative interim measure might include changes in the composition of the leadership staff (such as including a teacher from the school with strong leadership qualities).

Challenge 2: The time it takes for a district, CMO, or state to provide the necessary demographic, testing, principal, and school data for an evaluation is also subject to delay. We have found that there is often a lag of one year or more between the collection of data for a certain school year and providing those data to evaluators.

Recommendation for Policymakers and Evaluators: This potential lag should be kept in mind when planning timelines for the evaluation of school leadership improvement efforts.

Controlling for Student Characteristics

Using data at the student level is ideal when creating effect measures for evaluating principal improvement efforts. The use of student-level data allows the evaluator to control for various student-level factors that affect achievement. Controlling for these characteristics helps to isolate the effect of the effort being evaluated. Even so, limits remain in the ability of evaluation models to control for all of the potentially relevant factors using available data. Information known to be relevant to achievement outcomes, such as family background and composition, is not collected by schools and districts. Information that is collected, such as FARM status, may not adequately measure the underlying characteristics of interest.

Unobserved Characteristics

There are many student-level characteristics linked to achievement that evaluators could use to help control for various factors outside of the school. District data typically contain student information on race/ethnicity, gender, birth date, and test scores. Occasionally, these data contain other useful information, such as suspensions, courses taken, and individual class grades, that can be linked to student records.

Challenge: Students' performance in school is influenced by the actions educators take and the resources schools provide, as well as factors that are outside schools' control, such as family background and neighborhood conditions (see, for example, Hanushek, 1997; Ballou, Sanders, and Wright, 2004). A lack of data related to factors outside of school can impede the ability of an evaluation to isolate the contribution of the effort being evaluated from these other factors. Even if data were available on such characteristics as presence of siblings or a two-parent home, a multitude of unobserved factors could potentially influence the student's performance in school.

Recommendations for Evaluators:

1. *Fixed-effects model:* With two or more years of data, it is possible to separate out unobserved characteristics that do not change over time using student-level fixed-effects models. This type of model can control for characteristics that do not change even when those variables are not explicitly included in the model. However, any time-varying factors not included in the model will continue to influence estimates of principal effectiveness.
2. *Differencing/gain scores:* Other options to mitigate this problem include differencing or gain scores, which subtract a student's prior academic performance from his or her current performance, thus plausibly removing certain individual factors that may influence the student's test scores in all years. The methods are often not feasible for high schools, where test scores from a single point in time may represent an entire high school career. If a sufficient span of years is present in the data, it may be possible to subtract the high school student's eighth-grade test scores to create a gain score. However, gain scores do not necessarily remove the effects of outside-of-school factors and are not appropriate when the tests are not reasonably well-aligned across grade levels.

3. *Value-added measures:* A value-added measure for the principal effect could be used if multiple years of data are available.¹⁰ Such a measure helps control for various non-principal and nonschool factors that may affect student outcomes and isolates the contribution of the principal. However, even with multiple years of data, it may be that the student outcome data in a particular district have not been collected in a manner to support the creation of value-added measures. In this case, more weight should be placed on other methods of measuring performance. For example, in a situation in which the student outcome data are suspected of having widespread errors, other measures should be emphasized. There are also certain situations in which principal value-added measures cannot be constructed, such as for new schools (as they lack a prior history of student outcomes) and schools composed of nontested grades (kindergarten through second grade in many districts). High schools are also problematic for value-added measure creation, for reasons related to the frequency of testing mentioned above. Overall, value-added measures can solve some problems with unobserved factors at the student level but are not a panacea, due to the limitations mentioned above.
4. *Include background measures and prior outcomes:* Our recommendation in all cases is to acknowledge the possibility that unobserved student factors may impact an evaluation's ability to detect outcomes that can be traced to principals. However, the use of gain scores, value-added measures, or fixed-effects models can mitigate the effect of unobserved time-invariant student characteristics, as discussed above. Student background measures should be included in all models, as well as multiple prior outcome measures (such as past test scores), if available.

Observed Characteristics

Even for those student-level characteristics available in administrative data, complicating factors may limit the usefulness of those variables in an analytical model.

Challenge 1: A variable may be recorded in several districts but have different meanings in each district. For example, the student-level variable “graduation” in one district may flag students who receive a regular diploma; in another, “graduation” may indicate students who receive a regular or honors diploma or a GED through the district. A mobility marker for whether students switched schools may include transfer students entering from other districts or just those students who transferred between district schools. These anomalies can affect the attribution of outcome measures to the appropriate school, as well as the interpretation of results.

Recommendation for Evaluators: Evaluators should use caution with district-created markers and be sure they understand how these are created, especially if the effort covers multiple districts or CMOs. If possible, evaluators should obtain detailed entry and withdrawal codes from the district so that they can differentiate among student mobility, graduation, and dropout. The evaluator can then construct variables that reflect these different exit pathways.

Challenge 2: Data on student characteristics, especially those based on student or parent self-reports, may be incomplete or inconsistent. Families may respond in unexpected ways or

¹⁰ For further information on value-added measure construction for principals, see Lipscomb et al. (2010). The use of value-added measures to identify effects for teachers has been controversial because of instability in the estimates and challenges in assigning students to teachers for the purpose of constructing the measure. These concerns are lessened when developing value-added measures for principals, although there may still be instability issues.

refuse to respond to questions about race and ethnicity. For example, FARM status is often used as a marker for a student's socioeconomic status (SES).¹¹ However, students who are not identified as FARM-eligible could be either from households above the income threshold (the typical interpretation) or from households that qualify but did not apply for FARM. In Baltimore City Public Schools, close to 10,000 students were newly eligible for free or reduced-price meals in the 2009–2010 school year, due in part to more parents filling out the application for the program (Alonso, 2010). These issues make it challenging to interpret the data and could cause unexpected results.

Recommendation for Evaluators: Evaluators should carefully assess the observed characteristics in the data and consider the definition and how the variable was created and validated. Evaluators should also consider making use of student residence data that may be available. Using software to map census tract and block group onto the data will allow evaluators to link data from the American Community Survey (ACS) on neighborhood characteristics. These data contain information on average income in a census tract, which may be used in place of FARM as a measure of SES. Additionally, such variables as average adult education can also be found in the ACS and might serve as useful control variables.

Accounting for School Context

School context is an important mediator that is expected to influence the relationship between principal improvement efforts and student outcomes. School context may influence the actions that a principal takes in order to improve student outcomes, as well as the time it takes to realize gains. Many aspects of the school context are not observable, and even where differences in school context are observable, the manner in which the improvement effort is implemented may make it difficult or impossible for the analysis to control for them.

Determining Appropriate Comparison Schools

Ideally, when determining whether or not an effort designed to improve principal performance is successful, a randomized experiment would be conducted. This method would allow evaluators to compare two groups of principals where the only difference between the two groups was that one group was part of the improvement effort and the other group was not. This would allow evaluators to conclude that any differences in outcomes between the groups were a result of the improvement effort.

Challenge 1: Random assignment is typically not feasible in the context of principal improvement efforts. For the evaluation of the New Leaders program, random assignment of program principals to schools in a particular district was not an option, as each district has its own hiring policies and procedures. More generally, researchers are often confronted with stakeholder objections or administrative roadblocks when they recommend random assignment. Those who are responsible for implementing interventions typically object to arrange-

¹¹ It is important to note that FARM status is an imperfect proxy for SES. The Department of Agriculture's Child Nutrition Program's income eligibility guidelines for the time period of July 1, 2011, to June 30, 2012, state that those with an income 185 percent of the federal poverty level or lower are eligible for reduced-price meals, and those with an income 130 percent of the federal poverty level or lower are eligible for free meals (Federal Register, 2011). In addition, FARM status is not defined consistently for all grades and is not likely to accurately measure the economic resources of the household (Harwell and LeBeau, 2010).

ments in which the intervention only benefits some schools and students, especially when all of the authority to select the recipients of the intervention lies in the hands of the evaluators. Instead, researchers must choose a comparison group of principals not exposed to the improvement effort after the implementation of the improvement effort.

The New Leaders student outcome analysis compares the effects of the New Leaders principals with those of non–New Leaders principals. The fact that principals who participate in alternative training programs like New Leaders are often placed in atypical or unique school settings makes a simple comparison of the outcomes uninformative because it is likely that the difference in outcomes stems from a number of different factors, including school context. In New York City, for example, an evaluation of the NYC Leadership Academy’s Aspiring Principals Program (APP) found that these principals were more likely to be in low-performing schools that were notably different from comparison schools and had a downward academic trajectory prior to placement of the APP principal (Corcoran, Schwartz, and Weinstein, 2012).

Recommendation for Evaluators: To combat these differences in school context, our analyses include a wide variety of control variables:

- Student-level controls: FARM-eligible, race/ethnicity, gender, English-language proficiency, age, individualized education program (IEP) status
- School-level controls: Percentage FARM-eligible, average race/ethnicity, number of students, charter, magnet, percentage of students with IEP status
- Principal-level controls: Tenure, background information if available.

The availability of the control variables at both the student and school levels varies by district and CMO; we recommend that evaluators include appropriate control variables in analyses to mitigate effects of differential school contexts. In addition, efforts should be made, to the extent possible, to match schools with atypical or unique settings.

Recommendation for Policymakers: Policymakers are strongly cautioned to keep this issue in mind when assessing the effects of efforts targeting principals in these types of schools. It is important to understand and question the methods used to create a comparison group for evaluation purposes and to consider that there may be underlying reasons why one group may differ vastly from another.

Challenge 2: There are limitations to the control variables available for analysis. For example, district-provided and publicly available data are usually not up to date with the current school year. This can be problematic when the characteristics of schools are changing rapidly because of school closures, redistricting, or the opening of new schools.

Recommendation for Policymakers: Developing a data system that allows for more rapid updating of district data would allow the inclusion of start-up schools in evaluations and would better reflect changing school composition.

Challenge 3: Another area of school context is the leadership staff structure and the principal’s influence over leadership staff. Most of the case study schools had one or two formal leaders beyond the principal, such as assistant principals or school-based coaches, and principals had different opportunities to shape that team. In cases in which the principal has access to a leadership team, the composition and degree of principal influence over that team can influence the effect that principal has on student outcomes.

Recommendation for Evaluators: Where possible, evaluators should consider the size and structure of the school’s leadership team and the degree of principal influence over that

team. Because districts provide differing levels of principal autonomy over staffing, this influence is likely to vary by district. The issue of leadership teams should be mentioned in the evaluation report. Evaluators should provide guidance to readers to help them interpret the findings appropriately.

Developing Measures of School Context

School culture and teacher capacity play key roles in determining how quickly a principal can influence student outcomes. Improvements in these areas may be important indicators that principals are making progress that will eventually lead to improved student outcomes. Currently, systematic data on school culture and teacher capacity are not readily available. Through our work, we have explored a few possible metrics, such as the level of staff cohesiveness and level of teacher collaboration that shed useful light on school culture and teacher capacity. However, more work is needed to refine the measures used in this area.

Challenge: The context of the school in which the principal is placed may influence (either positively or negatively) the principal's impact on student achievement. Our case study work related to the New Leaders evaluation suggests that principals placed in low-performing schools with low teacher capacity experienced barriers to their leadership efforts. Student achievement level alone did not fully describe the school context.

Recommendations for Evaluators: When evaluating the effectiveness of efforts to improve school leadership, stakeholders need to take school context into consideration. In our work, we have developed some preliminary measures based on available administrative data to enable classification of all schools; we have also created richer measures of school context based on multiple measures. In addition to the measures we have been able to develop with our data, we have suggestions for additional measures using teacher-related metrics that can enrich the classification of schools further.

1. *Achievement groupings:* A principal's ability to influence school performance may depend on the state of the school at the time the principal enters a school. To account for school differences, our evaluation grouped schools into four categories based on prior achievement trajectories (Burkhauser et al., 2012). This allowed us to analyze outcomes while controlling for the achievement trajectory of the school upon new principal entry.
 - Start-up schools were schools that opened in the 2007–2008 school year.
 - Schools classified as above Adequate Yearly Progress (AYP) had met their AYP targets in both mathematics and reading in the prior school year.
 - Schools classified as below AYP and improving failed to meet mathematics and reading AYP targets in the 2006–2007 school year but were on an improvement trajectory.¹²
 - Schools classified as below AYP and not improving failed to meet math and reading AYP targets in the 2006–2007 school year and were not on an improvement trajectory.

¹² Schools were considered to be on an improvement trajectory if any of the following were met: (a) Their combined improvement in reading and math proficiency from the 2005–2006 school year to the 2006–2007 school year was at least 15 percent; (b) they were a start-up school in the 2004–2005 school year, and their combined improvement in reading and math proficiency from the 2004–2005 school year to the 2006–2007 school year was at least 20 percent; (c) their combined improvement in reading and math proficiency from the 2003–2004 school year to the 2006–2007 school year was at least 30 percent.

2. *Teacher-related metrics:* Better metrics on teacher capacity are needed to control for the context of the school and to assess interim progress. Most schools have a mix of novice and veteran teachers. The staff makeup presented different challenges to the principals in our study: For example, those with many novice teachers dealt with preventing burnout, while those with many veteran teachers worked to gain teacher buy-in and to foster a sense of urgency that continued improvement in student outcomes was necessary. However, teacher experience is only one dimension of teacher capacity. A number of other teacher-related metrics could be created using district data and used to assess a principal's progress toward student achievement. For example, teacher value-added metrics and other teacher quality metrics could be used in combination with teacher turnover data to assess whether teacher quality is improving over time and whether the principal is retaining effective teachers.

Measuring Principal Impact in Differing Contexts

Principals engage in a wide variety of tasks throughout the day and balance numerous responsibilities that include both the tactical (budgeting, staffing, transportation, etc.) and the strategic (human capital management, instructional leadership, creating a vision, etc.). For efforts that operate across multiple districts, or within both charter schools and regular district schools, the context in which these actions occur can vary dramatically. In particular, the degree of decisionmaking autonomy related to budget, curriculum, and school staffing issues can be high in one context and low in another. A principal who is unable to replace poorly performing teachers or lacks resources to offer necessary professional development to school staff has limited ability to improve teaching and thereby increase achievement test scores (Chiang, Lipscomb, and Gill, 2012).

Challenge: Various contextual factors, including the principal's level of autonomy over staffing decisions, can influence the magnitude of principal effects on student achievement. We have found that estimates of the effect are sensitive to changes in the schools that are included in the analysis. For example, we have found that the magnitude of the program effect varies by district. We also found evidence that the program had differential effects on charter schools, small schools, start-ups, and other nontraditional schools when compared with traditional public schools. This suggests that context has a potentially important influence on the pathway from the effort to an impact on student outcomes.

Recommendations for Evaluators:

1. *Understand context:* It is important to understand how these contextual factors may influence the effect of the effort and attempt to control for these factors in the evaluation. If it is not possible to control for these factors, or if the context constrains principals' actions in ways that make it highly unlikely that the effort will succeed, these conditions should be discussed and acknowledged as part of the evaluation.
2. *Type of school:* When evaluating a school leadership effort, separate analyses by type of school should be conducted in order to determine if the effort has more impact in certain types of schools, districts or CMOs, or regions. Alternatively, these differences could also be tested by including interaction terms in the statistical models, although the number of separate analyses may be limited by the number of schools in a particular category. In developing an evaluation strategy, the evaluator will need to anticipate the categories of schools across which effects would be expected to vary and take into

consideration the feasibility of separate analyses by school types in developing an evaluation strategy.

3. *Level of effect:* Those evaluating an effort spanning multiple districts or states should consider at what level—district, state, or regional—to estimate and report the effect of the program. One option is to estimate effects for various levels, which will provide a better understanding of the nature of the effect.

Recommendation for Policymakers: As context and grouping of schools matter, policymakers should assess the extent to which the results of other studies are generalizable to their school, district, or state. In adopting an effort that has been effective in another context, policymakers should pay particular attention to whether the new context has similar features and consider the potential implications of any differences.

Controlling for Principal Characteristics

Prior research suggests that it is particularly important to account for the level of experience a principal has leading a school, other leadership experience, and experience in the same school prior to becoming a principal when conducting an analysis of an effort to improve school leadership. Other factors, such as experience managing adults and prior teaching experience, may influence the magnitude of an effort's impact. Principals also bring a wide range of backgrounds and experiences to their schools, making the ability to isolate the effect of the effort beyond the effect of the particular principal difficult. To account for these characteristics and isolate the effect, such measures as school leadership experience could be included as control variables in the analysis.

Quality and Availability of Principal Tenure Data

A principal's tenure at a school may be related to student outcomes and is therefore an important variable to capture when evaluating efforts to improve student outcomes through principal improvement efforts (see, for example, Clark, Martorell, and Rockoff, 2009; Béteille, Kalogrides, and Loeb, 2011; Fuller, Baker, and Young, 2007; Dhuey and Smith, 2011; and Brewer, 1993).

Challenge: District-provided principal tenure data may not be accurate or may not capture the correct information. For the New Leaders evaluation, we obtained principal tenure data from both the district administrative records and from principals themselves. A comparison between the principal self-reports of school tenure and district data showed a concerning rate of discrepancy (21–42 percent across the districts). After a thorough audit of the district-provided principal tenure data, we discovered the following issues with the tenure data: errors in the data files from districts in one or more years, misunderstanding what the principal tenure data field is capturing, and lags in updating data when a new principal is assigned to a school or in removing an old principal from the file. Many of the districts warned us about the validity of the tenure data, especially from older files, because many districts have changed data systems, which makes it impossible to verify the earlier information.

After conducting our principal tenure audit in each district, we were able to correct the majority of issues with the data.¹³ Our method going forward is to collect a list of principals new to their schools in each school year for each district, setting tenure to one in the case of a new principal. In the case of a returning principal, we add one year to the tenure value for the prior year. With annual lists of new principals, we are able to directly control what our tenure variable represents.

Recommendations for Evaluators:

1. *Work with districts to interpret and improve tenure data:* A lesson from our experience is that researchers who work with district-provided principal tenure data should not take this data at face value and should work with districts and CMOs to gather the most accurate data possible. We recommend that evaluators work directly with districts to ensure that they understand what the district is capturing in the principal tenure variable and other data derived from files used for human resources functions. Human resources data files were originally created to support transactional functions (such as payroll processing and retirement eligibility determination). Many of these systems do not retain information in a way that can easily be used for evaluation research.
2. *Acknowledge limitations in tenure data:* When reviewing other evaluators' reports using principal tenure, evaluators should be aware that the tenure variable used in the report may be inaccurate. In their own work, evaluators should acknowledge any potential problems they encountered in recording principal tenure.

Recommendation for Policymakers: Districts and states should develop data capability to support their human capital activities, including principal recruitment, screening, and placement. While these data may originate from human resources databases, thought must be given to the data needs for evaluation purposes. State policymakers should consider encouraging systematic reporting of tenure variables (e.g., tenure as principal across districts, tenure in district as principal, and tenure in current school as principal) in districts and CMOs within their jurisdiction.

Variation in Principal Career and Training Paths

Principals may lead a school after a few years of teaching or after a lengthy career as an educator; the principal may or may not have professional experience outside of education. All principals have the basic requirements outlined by the state in which they serve, but professional and personal experience in many other areas varies greatly and may be related to performance as a principal. For example, serving as an assistant principal prior to becoming principal may improve performance as a principal, especially for those at an early stage in their career (Clark, Martorell, and Rockoff, 2009). States set minimum qualifications for certification as a principal; these typically include a certain number of years as a teacher in a K–12 classroom, a master's degree in school administration (or completion of another approved program), and passing a certification exam. Many states also require a valid teaching certification. The minimum required number of years as a teacher varies; it is two or three years in most states.

¹³ This audit involved helpful communication with district data personnel that enabled us to clarify what the tenure variables represented.

Challenge: There are numerous pathways to becoming a school leader, and it is often difficult to obtain the information on principal training and career trajectories needed to form appropriate comparison groups and include such information as a control in the evaluation. Information on training programs that a principal attended and the skills and training emphasized in these programs is also lacking, complicating comparisons among principals from different training programs. When evaluating a principal training program, these difficulties make it challenging to determine which aspects of principal training programs are the most effective. The ability to determine these most effective aspects would be helpful in designing targeted curricula for training programs and professional development courses.

Recommendation for Evaluators: Evaluators should use available principal information as controls in any analysis, as well as in forming comparison groups. When evaluating an effort, controls for principal characteristics and experience should be used; if these are not available, it will hamper the ability to isolate the effect of the effort on student outcomes. For smaller districts or evaluations of efforts that only target a small number of principals, it may not be possible to include principal characteristics as controls in a model, but these factors should still be considered and understood as having an effect on the impact of the effort.

Recommendation for Policymakers: From our research, we typically find that data available provide only a minimum amount of information about each principal's background and experience. Ideally, additional information that is usually present in the personnel files of principals but is rarely connected to school records would be useful for an evaluation. The following areas of data would be particularly useful:

- demographic data, such as gender, age, and race/ethnicity
- prior teaching experience
- experience as a principal outside the district. These data would make it possible to distinguish between novice principals and those who have prior experience but are new to the district.

Our recommendation is for districts and states to record additional principal information in a usable format and connect these data to student and school information in the data system. These data would be used when evaluating principal improvement efforts to control for background characteristics, to examine specific aspects of the effort in question, and to help form comparison groups.

Recommendation for CMOs: Principal data are often less available from CMOs than from districts and states, particularly in terms of principal experience; CMOs should evaluate and develop their data capability in this area.

Incorporating Multiple Measures

In concluding this chapter, we highlight a critical challenge that has come up in several of the sections for reasons having to do both with the pathway through which principals are expected to influence student outcomes and with the availability of data to measure the effect of the effort.

Challenge: Measuring only changes in student achievement does not reflect the breadth of a principal's influence at his or her school, provide indicators of progress, or account for con-

textual factors. Including additional information on principal practices provides a more well-rounded evaluation of the school leadership improvement effort's effect. Even using principal value-added measures and including control variables is not enough to reflect the contribution of the principalship; the data used to create value-added measures are often problematic.¹⁴

Recommendation for Evaluators: Multiple measures of principal practice, particularly in relation to organizational management, administrative tasks, and instructional tasks, should be incorporated into the evaluation if possible. Although indirect measures of practice, such as surveys, might be unavailable because of budgetary or logistical constraints, it is sometimes possible to obtain administrative data or artifacts that provide evidence regarding principals' practices. The choice of measures should be guided by the specific pathway through which a policy, practice, or program seeks to influence school leadership and student outcomes. For example, a program seeking to improve instructional leadership may wish to include measures of how the principal spends his or her time, measures of teacher perceptions on the principal's instructional leadership, and student outcomes.

¹⁴ For further information on value-added measures for principals, see Lipscomb et al. (2010).

Conclusion

Policymakers and funders are increasingly recognizing the central role that principals play in supporting and ensuring high-quality teaching in schools. This realization has piqued interest in efforts to improve the principalship and a desire to evaluate the effectiveness of such efforts.

Throughout our evaluation of the New Leaders program, we have grappled with a number of conceptual and practical challenges described in this report. We do not provide a neat answer or a single “best practice” that can be applied by all stakeholders in all circumstances. The evaluation of efforts that target principals presents unique challenges. The nature of those challenges and the opportunity to address them varies depending on the characteristics of the effort, such as how many schools it affects, the diversity of sites in which it is being implemented, and how long it has been in existence.

An overarching reality that policymakers and funders must come to terms with is that it will take time for efforts that target principals to show results. Whereas one might expect to see changes in student achievement from a curriculum intervention or even a targeted teacher professional development effort relatively quickly, efforts targeting principals typically influence student outcomes through the principal’s influence on teachers. In the case of a principal preparation program, these changes in student outcomes can easily take up to four years or more to appear from the time of the initiation of the program, depending on the outcome used. Even when an effort is comprehensive enough to support a rigorous outcomes-based evaluation, stakeholders will want to consider other measures that are available earlier in the effort’s life cycle.

Evaluators of efforts designed to affect the principalship face numerous challenges that make it difficult to isolate the effect of the effort. Overall, one of the most important recommendations is to use multiple measures when evaluating these efforts. Ideally, these multiple measures will include both qualitative and quantitative techniques and consist of both student outcomes and interim measures.

Policymakers also have a role to play in mitigating these challenges; they should assess the data situation and lay the groundwork for the collection and retention of data that will be needed for the purposes of evaluation at the start of an initiative. There may also be a role for state policymakers to encourage consistent and systematic reporting of core variables of interest by all districts.

The challenges that we have identified in this report can be addressed by improving the availability and quality of data, by choosing appropriate evaluation methods, and by appropriately interpreting the results of the evaluation. States, districts, and funders may find it challenging to adopt all of these recommendations in an environment of constrained resources but should bear in mind that time, money, and energy devoted to the evaluation of these types of

programs can help states and districts make the best use of available resources. In discussing common complications and recommending how to avoid, mitigate, or acknowledge them, we have identified a number of important considerations that can serve as a practical guide for evaluators and policymakers for the evaluation of school leadership improvement efforts.

References

- Alexander, Karl L., Doris R. Entwisle, and Carrie S. Horsey, "From First Grade Forward: Early Foundations of High School Dropout," *Sociology of Education*, Vol. 70, No. 2, 1997, pp. 87–107. As of July 12, 2012: <http://www.jstor.org/stable/2673158>
- Alonso, Andrés A., "You Make a Difference," Baltimore, Md.: Baltimore City Public Schools, March 1, 2010. As of July 12, 2012: http://www.baltimorecityschools.org/cms/lib/MD01001351/Centricity/Domain/87/2010_11_PDFs/FARM_2010_11.pdf
- Balfanz, Robert, Liza Herzog, and Douglas J. Mac Iver, "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions," *Educational Psychologist*, Vol. 42, No. 4, 2007, pp. 223–235.
- Ballou, Dale, William Sanders, and Paul Wright, "Controlling for Student Background in Value-Added Assessment of Teachers," *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, March 20, 2004, pp. 37–65. As of July 12, 2012: <http://jeb.sagepub.com/content/29/1/37.abstract>
- Béteille, Tara, Demetra Kalogrides, and Susanna Loeb, "Stepping Stones: Principal Career Paths and School Outcomes," Stanford, Calif.: Stanford University, 2011.
- Bonsteel, Alan, "California Acknowledges Middle-School Dropouts," *San Francisco Chronicle*, August 12, 2011, pp. A-8. As of July 12, 2012: <http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2011/08/11/EDHU1KMA31.DTL>
- Booker, Kevin, Tim R. Sass, Brian Gill, and Ron Zimmer, "The Effects of Charter High Schools on Educational Attainment," *Journal of Labor Economics*, Vol. 29, No. 2, 2011, pp. 377–415. As of July 12, 2012: <http://www.jstor.org/stable/10.1086/658089>
- Brewer, Dominic J., "Principals and Student Outcomes: Evidence from U.S. High Schools," *Economics of Education Review*, Vol. 12, No. 4, 1993, pp. 281–292.
- Burkhauser, Susan, Susan M. Gates, Laura S. Hamilton, and Gina Schuyler Ikemoto, *First-Year Principals in Urban School Districts: How Actions and Working Conditions Relate to Outcomes*, Santa Monica, Calif.: RAND Corporation, TR-1191-NLNS, 2012. As of July 12, 2012: http://www.rand.org/pubs/technical_reports/TR1191
- Chiang, Hanley, Stephen Lipscomb, and Brian Gill, "Is School Value-Added Indicative of Principal Quality?" paper presented at the Association for Education Finance and Policy 37th Annual Conference, Boston, Mass., the Association for Education Finance and Policy (AEFP), March 15–17, 2012.
- Clark, Damon, Paco Martorell, and Jonah E. Rockoff, *School Principals and School Performance*, Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, the Urban Institute, Working Paper 38, 2009.
- Corcoran, Sean P., Amy Ellen Schwartz, and Meryle Weinstein, "Training Your Own," *Educational Evaluation and Policy Analysis*, Vol. 34, No. 2, June 1, 2012, pp. 232–253. As of July 12, 2012: <http://epa.sagepub.com/content/34/2/232.abstract>
- Data Quality Campaign, home page, 2012. As of July 12, 2012: <http://www.dataqualitycampaign.org/>

Dhuey, Elizabeth, and Justin Smith, "How Important Are School Principals in the Production of Student Achievement?" AEFPP 36th Annual Conference Papers and Posters, 2011.

Federal Register, *Food and Nutrition Service: Child Nutrition Programs—Income Eligibility Guidelines*, Department of Agriculture, 2011. As of July 12, 2012:
<http://www.fns.usda.gov/cnd/governance/notices/iegs/IEGs11-12.pdf>

———, Vol. 73, No. 210, *Rules and Regulations*, October 29, 2008, pp. 64435–64513. As of July 22, 2012:
<http://www2.ed.gov/legislation/FedRegister/finrule/2008-4/102908a.html>

Fuller, Edward, Bruce D. Baker, and Michelle Young, "The Relationship Between Principal Characteristics, School-Level Teacher Quality and Turnover, and Student Achievement," Working Paper Series, Federal Reserve Bank of Atlanta, 2007.

Hamilton, Laura S., Brian M. Stecher, and Kun Yuan, "Standards-Based Accountability in the United States: Lessons Learned and Future Directions," *Education Inquiry*, Vol. 3, No. 2, 2012, pp. 149–170. As of July 12, 2012:
http://www.use.umu.se/digitalAssets/98/98129_inquiry_vol3_nr2_hamilton.pdf

Hanushek, Eric A., "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis*, Vol. 19, No. 2, June 20, 1997, pp. 141–164. As of July 12, 2012:
<http://epa.sagepub.com/content/19/2/141.abstract>

Harwell, Michael, and Brandon LeBeau, "Student Eligibility for a Free Lunch as an SES Measure in Education Research," *Educational Researcher*, Vol. 39, 2010, pp. 120–131.

Hauser, Robert M., and Judith Anderson Koenig, eds., *High School Dropout, Graduation, and Completion Rates: Better Data, Better Measures, Better Decisions*, The National Academies Press, 2011. As of July 12, 2012:
http://www.nap.edu/openbook.php?record_id=13035

Koretz, Daniel, *Measuring Up: What Educational Testing Really Tells Us*, Cambridge, Mass.: Harvard University Press, 2008.

Lipscomb, Stephen, Bing-ru Teh, Brian Gill, Hanley Chiang, and Antoniya Owens, *Teacher and Principal Value-Added: Research Findings and Implementation Practices*, Cambridge, Mass.: Mathematica Policy Research, Inc., 2010. As of July 12, 2012:
http://www.mathematica-mpr.com/publications/PDFs/education/teacherprin_valueadded.pdf

Martorell, Paco, Paul Heaton, Susan M. Gates, and Laura S. Hamilton, *Preliminary Findings from the New Leaders for New Schools Evaluation*, Santa Monica, Calif.: RAND Corporation, WR-739-NLNS, 2010. As of July 12, 2012:
http://www.rand.org/pubs/working_papers/WR739

New Leaders, home page, 2012. As of July 12, 2012:
<http://www.newleaders.org/>

Piro, Jody, Roger Wiemers, and Tammy Shutt, "Using Student Achievement Data in Teacher and Principal Evaluations: A Policy Study," *International Journal of Educational Leadership Preparation*, Vol. 6, No. 4, 2011. As of July 12, 2012:
<http://cnx.org/content/m41125/1.2/>

Rumberger, Russell W., "Dropping Out of Middle School: A Multilevel Analysis of Students and Schools," *American Educational Research Journal*, Vol. 32, No. 3, September 21, 1995, pp. 583–625. As of July 12, 2012:
<http://aer.sagepub.com/content/32/3/583.abstract>

Rumberger, Russell W., and Gregory J. Palardy, "Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance," *American Educational Research Journal*, Vol. 42, No. 1, March 20, 2005, pp. 3–42. As of July 12, 2012:
<http://aer.sagepub.com/content/42/1/3.abstract>

Rumberger, Russell W., and Scott L. Thomas, "The Distribution of Dropout and Turnover Rates Among Urban and Suburban High Schools," *Sociology of Education*, Vol. 73, No. 1, 2000, pp. 39–67. As of July 12, 2012:
<http://www.jstor.org/stable/2673198>

U.S. Department of Education, "Bringing Flexibility and Focus to Education Law," fact sheet, Washington, D.C., 2011. As of March 2, 2012:
http://www.whitehouse.gov/sites/default/files/fact_sheet_bringing_flexibility_and_focus_to_education_law_0.pdf

Wolf, Patrick J., *The Comprehensive Longitudinal Evaluation of the Milwaukee Parental Choice Program: Summary of Final Reports*, Fayetteville, Arkansas: University of Arkansas, 2012. As of July 12, 2012:
http://www.uaedreform.org/SCDP/Milwaukee_Eval/Report_36.pdf