

# Exploring the Causal Impact of the McREL Balanced Leadership Program on Leadership, Principal Efficacy, Instructional Climate, Educator Turnover, and Student Achievement

**Robin Jacob**

*University of Michigan*

**Roger Goddard**

*Ohio State University*

**Minjung Kim**

*University of South Carolina*

**Robert Miller**

*Texas A&M University*

**Yvonne Goddard**

*Ohio State University*

*This study uses a randomized design to assess the impact of the Balanced Leadership program on principal leadership, instructional climate, principal efficacy, staff turnover, and student achievement in a sample of rural northern Michigan schools. Participating principals report feeling more efficacious, using more effective leadership practices, and having a better instructional climate than control group principals. However, teacher reports indicate that the instructional climate of the schools did not change. Furthermore, we find no impact of the program on student achievement. There was an impact of the program on staff turnover, with principals and teachers in treatment schools significantly more likely to remain in the same school over the 3 years of the study than staff in control schools.*

**Keywords:** *randomized design, principal professional development, teacher turnover, principal turnover, principal leadership, principal efficacy*

RESEARCH clearly demonstrates a statistical link between the quality of school leadership, the instructional climate of the school, and student achievement (Branch, Hanushek, & Rivkin, 2012; Coelli & Green, 2012; Grissom, Kalogrides, & Loeb, 2012; Leithwood & Seashore Louis, 2012; Robinson, Lloyd, & Rowe, 2008; Waters, Marzano, & McNulty, 2003). Yet, despite this evidence, relatively little is known about how to most effectively develop school principals to achieve these outcomes. In particular, rigorous evaluations of programs

designed to provide professional development to help train and retain school leaders are almost nonexistent. A review by Camburn et al. in 2007 identified only two randomized control studies that evaluated the effectiveness of principal professional development programs; one that was conducted on 28 principals in 1970 and one conducted in 1986 that explored whether professional development was more effective when teachers and principals participated in the training workshops together. Since 2007, we know of only one randomized control trial of a principal

professional development program (Goldring, Camburn, Huff, Sebastian, & May, 2007). That study found little evidence for an impact of the program on principal behavior. However, the study suffered from a small sample size (48 principals in total) and low response rates—only 31 principals (13 treatment principals and 18 control principals) responded to the follow-up survey. We know of no studies that have explored the impact of principal professional development on the full range of potential outcomes, including leadership practices, school culture, staff turnover, and student achievement.

To add to the knowledge about the potential benefits of professional development for school principals, this study reports the results of a randomized control trial of one widely used professional development program for school leaders—McREL's Balanced Leadership® Professional Development (BLPD) Program—and assesses its impact on principal efficacy, leadership practice, the instructional climate of the school, staff turnover, and student achievement in a sample of rural northern Michigan schools.

The BLPD program is designed to provide research-based guidance to principals to help them enhance their effectiveness and improve student achievement. It focuses on teaching school leaders 21 key leadership responsibilities, such as visibility, order, and discipline, that have been shown to be associated with student achievement. The program involves a series of 10, two-day, cohort-based professional development sessions and has been in high demand throughout the United States and abroad since the program's inception. More than 20,000 school leaders in the United States and worldwide have participated in McREL's BLPD training through on-site sessions. In addition, the book *School Leadership That Works: From Research to Results* (Marzano, Waters, & McNulty, 2005), which outlines the BLPD approach, has sold more than 141,000 copies since its publication in September 2005.

Although McREL's own internal evaluations of the professional development program suggest that it is effective, to date no rigorous scientific evaluation has been conducted on the impact of the program. This study rigorously tests the impact of the program on a variety of outcomes

by randomly assigning 126 principals in Michigan's rural schools to either receive the BLPD Program or to a "business as usual" control group that followed standard district approaches to school improvement.

The article proceeds as follows. We begin by describing the BLPD program and its development. We then describe our hypothesized causal model and review the relevant literature on school leadership. We then describe the design of our study and the analyses we conducted. Finally, we present the results and discuss their implications.

### **The BLPD Program**

The McREL Balanced Leadership (BL) Framework is based on a meta-analytic work conducted by Marzano and colleagues. That meta-analysis (Waters et al., 2003) found a statistically significant relationship between school-level leadership and student achievement, and identified 21 individual leadership responsibilities (e.g., monitor instruction, involvement in curriculum, instruction and assessment, etc.) with statistically significant relationships to student achievement. These 21 responsibilities and the associated practices used by principals to fulfill these responsibilities (see the online appendix, available at <http://epa.sagepub.com/supplemental>) constitute the key content of the BLPD program. In addition, the program places a heavy emphasis on helping principals understand the focus of the change that is required to achieve certain outcomes and the complexity, or magnitude of the required changes. Finally, the program emphasizes the importance of creating a purposeful community to focus organizational resources on agreed upon goals. The responsibilities on which the BLPD program focuses are consistent with other recent professional development initiatives, like that of the Wallace Foundation, which emphasizes five key practices for effective principals: shaping a vision of academic success for all students; creating a climate hospitable to education; cultivating leadership in others; improving instruction; and managing people, data, and processes to foster school improvement (Wallace Foundation, 2013) and is consistent with the content of most educational administration programs as well as the Interstate

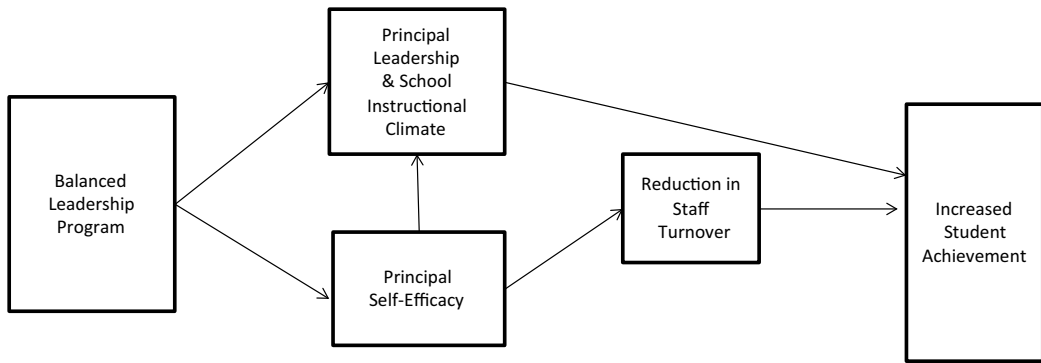


FIGURE 1. *Hypothesized pathways of influence connecting the balanced leadership program to student achievement.*

School Leaders Licensure Consortium (ISLLC) standards (ISLLC, 2008).

The program is organized to deliver four different types of knowledge deemed important for improving practices: *declarative* (knowing *what* to do), *procedural* (knowing *how* to do it), *experiential* (knowing *why* it is important), and *contextual* (knowing *when* to do it). Throughout, the sessions use a case methodology approach in which participants extend and refine their knowledge and understanding of the BL Framework by reading, reflecting upon, and discussing real-life cases of principals facing problems of practice, including applications to their own schools. Principals continually receive and share feedback from facilitators and peers.

The BLPD Consortia are facilitated by an experienced team of full-time training consultants, with extensive school-level leadership experience. To be authorized to facilitate the BLPD Consortia, staff members must undergo a quality assurance program, which consists of learning the content of the program with the guidance of an experienced BLPD facilitator, who also oversees practice sessions at McREL and then in the field.

### Hypothesized Causal Model and Literature Review

This study was designed to assess the impact of the BLPD program on principals' leadership practices, the instructional climate of the school, principal efficacy, staff turnover, and student achievement. Our hypothesized causal model is

displayed in Figure 1. Because the stated goal of the BLPD program is to improve the leadership skills of principals by helping them to better understand and implement the 21 key leadership responsibilities, manage change, and build purposeful communities, we hypothesized that the BLPD program would influence participating principals' leadership practices, including their knowledge of curriculum and instruction, their ability to shield staff from unwanted intrusions, their ability to lead and manage change and to provide intellectual stimulation for their staff. Relatedly, we hypothesized that program participation would lead to an improved instructional climate characterized by greater collaboration among staff, stronger norms for differentiated instruction to meet students' needs, and a general sense of trust and collective efficacy for achieving academic goals, elements which prior research has identified as key to creating a purposeful community (Goddard, Hoy, & Woolfolk Hoy, 2000, 2004; Goddard, Salloum, & Berebitsky, 2009; Goddard, Tschannen-Moran, & Hoy, 2001).

We also posited that participating in the BLPD program would improve principals' overall sense of efficacy. The construct of principal efficacy is grounded in Bandura's (1977) social cognitive theory, and refers to the degree to which principals believe that they can lead future improvements in instruction in their schools, and thus is expected to strongly influence the effort and persistence with which principals pursue instructional improvement in their schools. Because the BLPD program provides extensive peer-to-peer

interaction around effective practice and case study with application to participants' schools, we expected it could provide the types of enactive, vicarious, and socially persuasive experiences theorized by Bandura as key to the formation of efficacy beliefs. Alternatively, principal self-efficacy has been shown to be positively related to a variety of aspects of principal leadership, including setting directions, developing people, managing the instructional program, and classroom conditions (Leithwood & Jantzi, 2008). Therefore, increases in self-efficacy might also result in better leadership practices and an improved instructional climate in the school.

Leadership skills, the instructional climate of the school, and principal self-efficacy beliefs are each hypothesized to lead to reductions in staff turnover. Research by DeAngelis and White (2011) suggests that principals who leave their positions may not perceive themselves to be well-suited for their roles. Thus, if professional development increases a principal's knowledge and skills to a degree to which a better fit between background and role is perceived, the principal may be less likely to leave. Furthermore, efficacy beliefs have been hypothesized to influence the level of effort people put into activities and how long they will persist in those activities when faced with difficulty or failure (Bandura, 1977). Thus, principals in a hard to manage school might be less likely to leave if they have strong efficacy beliefs. This notion is supported by the research by Branch et al. (2012) who find that the most effective principals tended to remain in their positions even when they worked in the highest poverty schools.

Similarly, research shows that teachers are more likely to stay in a school where they feel the principal is a strong instructional leader (Allensworth, 2012; Boyd et al., 2011; Grissom, 2011; Ladd, 2011). Thus, improvements in the instructional climate of the school and the leadership skills of the principal should lead to a reduction in teacher turnover as well. Prior research has also demonstrated that teacher turnover is related to principal turnover; teachers are more likely to leave a school if and when their principal does (Béteille, Kalogrides, & Loeb, 2012). Therefore, if the program caused a reduction in principal turnover, it might also lead to a reduction in teacher turnover.

As research demonstrates an association between both principal leadership and the instructional climate of the school and student achievement (Branch et al., 2012; Bryk, Bender Sebring, Allensworth, Luppescu, & Easton, 2010; Coelli & Green, 2012; Goddard et al., 2001; Grissom et al., 2012; Robinson et al., 2008; Waters et al., 2003), improvements in both leadership practices and reductions in staff turnover are hypothesized to lead to increases in student achievement. Similarly, numerous studies have demonstrated a negative association between both principal turnover and teacher turnover and student achievement (Béteille et al., 2012; Miller, 2013; Ronfeldt, Loeb, & Wyckoff, 2013; Weinstein, Schwartz, Jacobowitz, Ely, & Landon, 2009).

Based on this causal model and our review of the literature, we identified the following research questions that our study design would allow us to address causally:

**Research Question 1:** What is the impact of the BLPD program on principals' leadership practices and the school's instructional climate (i.e., school climate, norms for teacher collaboration, norms for differentiated instruction)?

**Research Question 2:** What is the impact of the BLPD program on principal's efficacy beliefs?

**Research Question 3:** What is the impact of the BLPD program on teacher and principal turnover?

**Research Question 4:** What is the impact of the BLPD program on student achievement?

## **Method**

We begin by describing how the study was designed to test the research questions outlined above. Following that, we describe our sample, discuss attrition analysis, and present our approach to data collection and analysis of psychometric properties for survey measures.

### *Study Design*

To estimate the impact of the BLPD program, we recruited and then randomly assigned 126 schools, half to a treatment group in which the

principals participated in the BLPD program, and half to a “business as usual” control group. Recruitment was conducted by contacting superintendents who then signed a letter of agreement allowing their principals to participate. Principals were then contacted and invited to participate in the program. Principals assigned to the BLPD program received the professional development free of charge. Principals in the control group received whatever professional development their district or the state already supplied during the study period and were offered the chance to participate in an abbreviated version of the program for free at the conclusion of the study.

Principals in the treatment group were assigned based on geographic location to participate in one of two BLPD Consortia. Two consortia were planned to make travel to common locations feasible and to make the number of cohort participants typical of the size McREL consultants typically serve during consortia. The consultants selected by McREL to deliver each of these consortia were experienced facilitators with a strong record of performance. Each consortium was provided the same treatment delivered through 10, two-day BLPD training sessions from January 2009 to November 2010.

To conduct random assignment, we stratified the 126 schools in our sample based on U.S. Census Bureau locale code and socioeconomic status (SES).<sup>1</sup> We dichotomized the SES indicator with a median split classifying eligible schools into high and low poverty groups based on the percentage of students receiving free or reduced price lunch (FRPL). We then organized the sample into 12 cells (six locale codes  $\times$  two poverty categories) and assigned each school a random number. Next, we sorted each cell from lowest to highest by the random number. The first half of the schools in each sorted list was assigned to the treatment group; the second half was assigned to the control group.

### *Sample*

The original sample of 126 schools was drawn from the population of Michigan’s rural schools in the northern geographic third of the state. Many of these schools were located in Michigan’s Upper Peninsula and the rest were rural schools in geographically isolated northern regions of

Michigan’s Lower Peninsula. Both areas had weak local economies and had experienced double-digit unemployment rates immediately before and during the years of the study. These schools served disproportionately high proportions of students receiving an FRPL and had large proportions of students performing below state expectations on the Michigan Educational Assessment Program (MEAP), the annual state assessment employed to determine whether schools meet the adequate yearly progress requirements of No Child Left Behind (NCLB). Furthermore, given the severe economic status of the state, they also suffered from unusually constrained resources for high-quality professional development for school leaders.

Only schools that were (a) public (including magnet and charter), (b) served Grades 3 to 5 inclusively (e.g., K–8, PreK–5, 3–5, K–12), and (c) classified by the National Center for Education Statistics (NCES) rural or town locale code designations between 31 and 43 were eligible to participate in the study.<sup>2</sup>

Descriptive statistics for the treatment and control schools in the sample are shown in Table 1. The data demonstrate that the schools in our sample were small (around 300 students total compared with an average of around 400 per school statewide), poor (on average 47% of the students are eligible for FRPL), and mostly White (approximately 10% of the students in these schools were identified as an ethnicity other than White). Students in these schools scored right around the state average on the MEAP in both reading and math. Importantly, in no case was there a statistically significant difference between schools in the treatment and control groups on baseline demographic characteristics of the schools or the principals.

The schools that participated in the study were unique in that they were from the rural northern region of Michigan, when large randomized trials of educational interventions tend to focus on schools in larger urban areas. Because they are often not invited to participate in such studies, the schools in this sample were unusually eager to participate, as is reflected in the high survey response rates and high rates of attendance at the professional development sessions. In addition, the remote location and the economic status of the state during this time meant that there were

TABLE 1

*Descriptive Statistics for Treatment and Control Schools and Principals (N = 126)*

Variable	Treatment ( <i>n</i> = 62)	Control ( <i>n</i> = 64)	<i>t</i> value <sup>a</sup>	<i>df</i>	<i>p</i> value
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )			
% subsidized lunch	0.48 (0.14)	0.48 (0.13)	−0.28	124	.78
% minority students	0.10 (0.16)	0.12 (0.14)	−0.53	124	.60
Pupil–teacher ratio	17.02 (0.71)	16.65 (0.41)	−0.44	124	.66
School size	313.79 (153.59)	311.28 (164.27)	0.09	124	.93
2007 fourth-grade reading	430.09 (13.11)	429.00 (15.19)	0.43	123	.67
2007 fourth-grade math	427.77 (9.50)	424.61 (10.38)	1.77	123	.79
	Treatment (%)	Control (%)	$\chi^2$	<i>df</i>	<i>p</i> value
Locale code <sup>b</sup>			1.96	6	.923
Suburb, small	0.00	1.56			
Town, fringe	1.61	1.56			
Town, distant	1.61	1.56			
Town, remote	24.19	21.88			
Rural, fringe	11.29	7.81			
Rural, distant	29.03	29.69			
Rural, remote	32.30	35.90			
Principal characteristics	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>t</i> value <sup>a</sup>	<i>df</i>	<i>p</i> value
% female	0.47 (0.50)	0.39 (0.49)	0.90	122	.37
% master's degree or higher	0.81 (0.40)	0.74 (0.44)	0.85	122	.39
% White	1.00 (0.00)	1.00 (0.00)	NA		

<sup>a</sup>Fisher's exact test.<sup>b</sup>Likelihood ratio test.

few other professional development opportunities available to these principals. Thus, this sample not only informs the field regarding the impact of professional development in a rural context, which is virtually nonexistent, but also represents perhaps a best case scenario in which to test the effectiveness of a professional development program of this type.

#### *Attrition Analysis*

We were able to obtain student achievement data for 119 of the 126 schools that were initially recruited and randomly assigned as part of the study. Seven schools for whom we were unable to obtain student achievement data either closed or merged during the study period. Three of these were treatment schools and four were control. This yields an overall attrition rate of 5.6% and a

differential attrition rate of less than 2%; attrition rates that are unlikely to lead to bias (What Works Clearinghouse, 2011). Similarly, we were able to obtain information on teacher and principal turnover from 122 of the 126 schools.<sup>3</sup> Again, this small amount of attrition is unlikely to introduce bias into our results.

However, as shown in Figure 2, we were not able to obtain survey data from all participants at each wave of the study. Between random assignment and the baseline data collection, a total of 31 schools declined further participation in the study, 20 treatment schools and 11 control schools. Some of these schools actively declined to participate in the study after they were notified of the results of the random assignment. Others simply failed to return baseline surveys. Between baseline and the third year of data collection, we lost an additional treatment school and three control

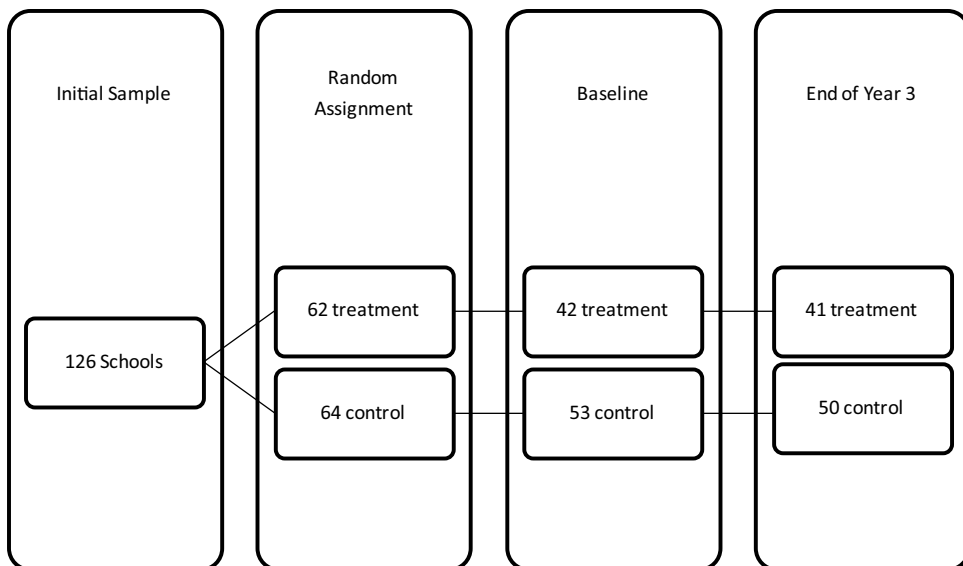


FIGURE 2. *Pattern of attrition for survey data.*

schools. Thus, for the teacher survey-based outcomes, our overall attrition rate was 28% with a differential attrition rate of 12%. This introduces the possibility that our survey findings may be biased (What Works Clearinghouse, 2011).

To assess whether attrition for survey completion was systematic, we compared the schools on a range of school and principal demographic characteristics and achievement in fourth-grade reading and math at the time of sample construction (spring/summer 2007) for both the baseline and final outcome survey completion rates. At baseline, we conducted *t* tests comparing the 95 participating schools with the 31 attriting schools. The results, shown in Table 2, do not indicate any systematic attrition for either treatment or control schools. There were no statistically significant differences ( $p \leq .05$ ) between attritors and nonattritors. We also compared the participating control group of 53 schools to the 11 control school attritors and to the initial total group of 64, and the 42 participating treatment schools to the 20 attriting treatment schools with similar results.

For the final outcome survey data collected 2 years later, there were four additional schools that did not provide teacher or principal survey data and three additional schools that provided teacher data but no principal data. Therefore, we performed the same analyses for the 91 participating

schools with teacher data and the 88 schools with principal data and found no statistically significant differences in either case. Similar analyses comparing the attritors and nonattritors in the full sample also found no differences. Although we did not find evidence of systematic attrition for survey completion, given the 15% differential attrition rate between the treatment and control groups for survey data, we chose to control for all available demographic characteristics in all subsequent analyses of treatment impacts.

### *Data Collection*

To obtain information about the impact of the BLPD program on principal efficacy, school leadership and the school's instructional climate (e.g., school climate and school norms for teacher collaboration and differentiated instruction), we rely on surveys administered to all principals and teachers in both the treatment and control schools at two points: (a) between November and December, 2008, just prior to the first intervention training (baseline) and (b) approximately 2 years later, after all intervention training was delivered (final outcome). Response rates for both principals and teachers were high for both waves of data collection. Specifically, at baseline, we received survey responses from 95% of all principals and 92% of all teachers, whereas

TABLE 2

*Attrition Analyses (N = 126)*

Variable	Attritors ( <i>n</i> = 31)	Nonattritors ( <i>n</i> = 95)	<i>t</i> value <sup>a</sup>	<i>df</i>	<i>p</i> value
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )			
% subsidized lunch	0.48 (0.13)	0.48 (0.14)	−0.19	124	.85
% minority students	0.13 (0.19)	0.10 (0.14)	−0.95	124	.34
School size	321 (190)	310 (148)	−0.35	124	.73
Teacher–pupil ratio	17.48 (7.31)	16.6 (2.24)	−0.90	124	.37
2007 fourth-grade reading	426.20 (2.08)	426.17 (0.98)	−0.01	123	.98
2007 fourth-grade math	429.45 (2.38)	429.57 (1.50)	0.04	123	.96
	Attritors (%)	Nonattritors (%)	$\chi^2$	<i>df</i>	<i>p</i> value
Locale code <sup>b</sup>			8.77	5	.19
Town, fringe	0	1.05			
Town, distant	6.45	2.11			
Town, remote	25.81	22.11			
Rural, fringe	12.90	8.42			
Rural, distant	29.03	29.47			
Rural, remote	25.81	36.84			
Principal characteristics	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>t</i> value <sup>a</sup>	<i>df</i>	<i>p</i> value
% female	0.42 (0.50)	0.43 (0.50)	0.10	122	.92
% master’s degree or higher	0.68 (0.48)	0.80 (0.40)	1.49	122	.15
% White	1.00 (0.00)	1.00 (0.00)	NA		

<sup>a</sup>Fisher’s exact test.

<sup>b</sup>Likelihood ratio test.

final outcome data were received from 97% of principals and 91% of teachers.

We surveyed all principals in the study schools both times regardless of how long they had been in the school and we did not follow, and did not offer professional development to principals if they left study schools. Similarly, we collected teacher survey and student achievement data only for those teachers and students who were in study schools at the time of data collection. Thus, our estimates of program impact reflect those one might expect given the ecological reality of principal, teacher, and student mobility over the course of a multi-year training program, which itself may have been impacted by the program.

We collected student-level MEAP scores from fall 2009 and 2011. These restricted-use assessment scores, obtained from the state, served as our achievement measures. Figure 3 illustrates how the data collection intersected with the

implementation of the program. Teacher and principal turnover data were also obtained from the state.

### *Measures*

This section describes the survey-based measures we employed, as well as the administrative turnover data and student achievement data obtained from the state.

*Survey Data.* The surveys were designed to assess the 21 leadership responsibilities identified by McREL as key factors in a principal’s ability to improve achievement and covered a range of additional topics including school norms for teachers’ collaborative and instructional practices, the level of trust among the staff, and principals’ sense of efficacy for leading instructional improvement. In cases where



	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug
Year 1 (08-09)		MEAP	Survey W1			BL1	BL2	BL3				BL4
Year 2 (09-10)		BL5	BL5/BL6			BL6 BL7		BL7 BL8		BL8		BL9
Year 3 (10-11)		BL10 MEAP			Survey W3							

FIGURE 3. Study timeline.

Note. One of the professional development (PD) locations had to reschedule PD Session 5 due to weather, thus the two locations received PD on a slightly different schedule in the Winter 2009–2010. MEAP = Michigan Educational Assessment Program testing; W1 = Wave 1; W2 = Wave 2; W3 = Wave 3; BL = Balanced Leadership PD Session.

previous researchers had developed psychometrically sound survey items to assess the constructs of interest, we used existing survey measures. If established items were not available, we constructed our own. For the most part, principals and teachers were asked to respond to parallel sets of questions. For example, principals were asked to respond to the question “I successfully protect teachers from undue distractions from their teaching,” while teachers were asked to respond to the question “The principal protects teachers from undue distractions from their teaching.” For most items, both principals and teachers were asked to respond on a 6-point Likert-type scale (*strongly disagree* to *strongly agree*). In few cases, respondents were asked to report on frequency of certain types of collaborative behaviors among teachers.

We employed exploratory and confirmatory factor analysis at baseline to assess the psychometric properties of our measures and performed the same analyses for final outcome data, which confirmed the factor structures we identified at baseline. Because principal and teacher surveys contained, for the most part, identical items, we were able to construct comparable measures for both principals and teachers. We initially identified 18 separate factors related to the concepts of interest. However, to reduce the problem of multiple comparisons, we drew on support from the theoretical and empirical literature to combine most sub-factors into three aggregate measures tapping principal leadership, school climate, and school-wide collaboration for instructional improvement. Seven of the original 18 measures (e.g., change, visibility, trust in principal) loaded onto a single measure of principal leadership. The model fit

statistics for this confirmatory factor analysis supports a one-dimensional leadership measure. Similarly, seven measures supported a single factor related to school climate with adequate model fit statistics. For collaboration, a single latent construct of collaboration comprised of four sub-constructs (i.e., teachers’ collaboration on instructional policy, frequency of collaboration on instruction, formal collaboration, and informal collaboration) was developed. The model fit statistics supported the factor model.

Because confirmatory factor analysis did not support the inclusion of subscales for principal efficacy (based on principal surveys only) or school norms for differentiated instruction in these three aggregate factors, they were retained as separate outcomes, yielding a total of five outcome measures constructed from principal survey data and four from teacher survey data.<sup>4</sup>

These five aggregate factors—principal efficacy (principal survey only), principal leadership, school-wide collaboration, norms for differentiated instruction, and school climate—serve as our survey-based measures of principal leadership, principal efficacy, and instructional climate. Observed factor scores, combining all of the items for each subscale were used in analyses. Descriptions of each measure and its statistical properties are shown in Table 3.

*Administrative Data.* From a restricted-use data file obtained from the Michigan Department of Education, we were able to construct measures of principal and teacher turnover during the course of the study. Using the unique ID numbers assigned to the teachers and principals in our sample, we were able to determine which teachers and principals who were employed in a

TABLE 3  
*Description of Survey Measures*

Measure	Description	Sample items		Cronbach's $\alpha$		Baseline $M$ ( $SD$ ), [minimum, maximum]	
		$t$	$p$	$t$	$p$	$t$	$p$
Principal leadership	Overall measure of a principal's ability to lead the school.	.987	.944	4.45 (1.01), [1, 6]	4.70 (0.52), [3.3, 5.8]		
Principal efficacy	Measure of how strongly a principal believes in his or her ability to affect change.	NA	.926	NA	4.49 (0.79), [2.4, 6]		
Collaboration	The frequency with which school staff collaborates, formally and informally and around topics related to instructional practice.	.881	.883	4.32 (0.77), [1.7, 6]	4.48 (0.67), [2.7, 6]		
Differentiated instruction	The degree to which teachers design instruction to meet the various needs, and learning strengths of their students.	.893	.914	4.72 (0.82), [1.4, 6]	4.51 (0.79), [2.2, 6]		
School climate	Measure of trust and sense of collective responsibility around achieving the school's academic goals.	.954	.930	4.66 (0.63), [2.3, 6]	4.77 (0.50), [3.4, 5.8]		

*Note.* All scales use a Likert-type response (range = 1–6).

participating school during the baseline year were in the same school at the end of the third year of the study. We used this information to construct measures of principal and teacher turnover.

*Student Achievement Data.* We used scale scores on the MEAP to assess the impact of the BLPD program on student achievement. The MEAP is administered each October to assess students' mathematics and English language arts achievement in Grades 3 through 8. We used student scores in treatment and control schools in October 2008 as baseline (prior to intervention) achievement measures. Student-level achievement scores from October 2011 were used to measure the impact of the program after implementation. The student achievement data file also included basic demographic characteristics of the students in the sample, which were used as covariates in our estimation models.

### Analysis

As described above, this study is designed to answer the following questions about differences between treatment and control schools regarding the impact of the BLPD program: (a) What is the impact of the BLPD program on principals' leadership practices and the school's instructional climate (i.e., school climate, norms for teacher collaboration, norms for differentiated instruction)? (b) What is the impact of the BLPD program on principal's efficacy beliefs? (c) What is the impact of the BLPD program on teacher and principal turnover? (d) What is the impact of the BLPD program on student achievement?

The following section describes our analytic approach to answering these questions.

*General Approach.* To estimate the impact of participation in the BLPD program on student achievement, we estimated the following hierarchical (or mixed) model (Raudenbush & Bryk, 2002):

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\text{Gender}_{1j} + \beta_{2j}\text{Minority}_{ij} + \beta_{3j}\text{SES} + r_{ij},$$

where  $Y_{ij}$  is the MEAP score for student  $i$  in school  $j$ ,  $\beta_{0j}$  is the regression-adjusted mean

value of student achievement for school  $j$ ,  $\beta_{1j}$ ,  $\beta_{2j}$ , and  $\beta_{3j}$  are the values of the coefficients on student-level covariates Gender, Minority Status, and SES, respectively, and  $r_{ij}$  is the residual error for student  $i$  in school  $j$ , which is assumed to be independently and identically distributed.

Level 2:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + \gamma_{01}T_j + \gamma_{02}\text{BaselineMEAP}_j \\ & + \gamma_{03}\text{SchoolSize}_j + \gamma_{04}\% \text{Minority}_j \\ & + \gamma_{05}\% \text{FreeLunch}_j + \mu_j, \end{aligned}$$

where  $\gamma_{00}$  is the grand mean of the regression-adjusted MEAP score for the average control school,  $\gamma_{01}$  is the estimated impact of treatment,  $T_j$  is equal to one for treatment schools, and zero for control schools,  $\gamma_{02}$ ,  $\gamma_{03}$ ,  $\gamma_{04}$ , and  $\gamma_{05}$  are the values of the coefficient on school-level average baseline achievement, school size, % minority, and % free lunch eligible, respectively, and  $\mu$  is the residual error for school  $j$ , which is assumed to be independently and identically distributed.<sup>5</sup>

Similar models were used to estimate impacts on teacher outcomes (a two-level hierarchical linear modeling [HLM] with teachers nested within schools) and principal outcomes (an ordinary least squares [OLS] analysis). Basic demographic characteristics (gender and highest level of education) of the teachers and principals were included in these models. Although in theory, in a random assignment study, unbiased estimates of program impact can be obtained by simply comparing mean differences in outcomes, we have included covariates in all of our models as a way to account for possible chance differences between treatment and comparison schools, to account for differential attrition in the case of the teacher and principal surveys, and as a way to decrease the standard error in the estimate of program impacts, thus increasing the power of the study.

*Missing Data.* As in any study, we experienced some degree of missing data. For principal survey data, we only included schools in which we had survey data from principals at both baseline and follow-up. This resulted in a total sample size of 88 principals. As principal turnover is endogenous to the treatment, we used the baseline data from the principal in the school at the start of the study, regardless of whether or not that same principal was present in the third year

of the study. In this way, if, for example, the treatment induced weaker principals to leave their jobs and those principals were replaced by stronger principals, the change would be captured in our analyses.

For the teacher survey data, we included in our outcome measure any teacher who was present in the school and responded to the survey in the third and final wave of the survey. The baseline measure was the average score for all teachers who were in the school at baseline. Again, this allowed us to capture any treatment-induced changes in the composition of teachers in the school. This resulted in a total sample size of 1,546 teachers in 91 schools. There were three schools in which we received survey responses from teachers, but not principals.

For student achievement, we obtained data from 119 of the 126 schools that were initially randomly assigned, although not all schools served all grades in each year of interest. The baseline measure was the average achievement of the students in that grade at baseline. Student-level covariates were provided to us as part of the student achievement file and school-level covariates, other than the baseline measure of achievement, were obtained through the Common Core of Data and there were no missing values.

*Intent to Treat Versus Treatment on the Treated.* As already mentioned, offering the BLPD program to principals in a particular group of schools did not guarantee that principals actually participated in the program. Some principals left study schools or simply decided not to attend some or all of the BLPD training sessions, while other treatment school principals never participated in any after learning of their random assignment status. By estimating impacts on student achievement and turnover in all the schools included in the sample when the study began, regardless of whether or not the principals actually participated in the BLPD program, we were able to obtain unbiased estimates of the “intent to treat” or the impact of *offering* the program to principals. The “intent to treat” estimate answers an important policy relevant question, namely, the impact of implementing the BLPD program in a group of schools across multiple districts given a certain amount of principal turnover or noncompliance.

However, these estimates cannot tell us what the impact of the program would have been if all principals had participated fully in the BLPD training, which is also an important policy relevant question. We therefore also estimate what is often referred to as the effect of the “treatment on the treated,” using a correction suggested by Bloom (1984), in which the intent to treat estimate is scaled up by the proportion of treatment principals who actually participated in the program. Such analyses provide better estimates of the effect of the treatment in schools where the principal participated in the BLPD program. Our estimates of treatment on the treated are based on the percentage of principals in the treatment group who ever attended a BLPD session (66% of the principals originally assigned to the treatment).

## **Results**

Before delving into the impact findings, we set the context for understanding our results by describing the fidelity of implementation and also the counterfactual condition (what professional development experiences the principals in the control group received) in the study schools. We then describe the findings from our surveys as well as administrative and assessment data.

### *Fidelity of Implementation*

First, we briefly summarize the findings from our qualitative investigation of implementation fidelity. Findings indicated that facilitators all had prior school leadership experience, were well trained by experienced facilitators who provided feedback on content and presentation, and were involved in a process of continuous improvement. Observation of training sessions by outside observers indicated the facilitators were highly skilled, possessed substantial expertise with the BLPD program content, and demonstrated consistency in adhering to the intervention training protocols. Principals assigned to the treatment group who agreed to participate after random assignment, attended 74.4% of the 20 BLPD program sessions (10 two-day workshops), on average, and more than half attended 90% or more of the sessions. Satisfaction with the professional development was high, with

87% of the participants rating the sessions as “very good overall” on surveys administered at the end of each session.

### *Counterfactual Condition*

Although a specific treatment was not designed for the comparison schools, principals in these schools received whatever professional development their school or district was currently offering. We asked questions on surveys to ascertain whether there was any contamination of the control group principals—that is, whether control group principals had been exposed to other programs similar to the BLPD program or sought out BLPD materials on their own. In particular, we asked about two programs being offered to principals in Michigan during the course of the study, which included content related to that of the BLPD program. We found that 12% of the principals in the control group had attended a program with content similar to Balanced Leadership—The Michigan Leadership Improvement Framework Endorsement (MI-LIFE) program. We also asked how many of the principals had read *School Leadership That Works*, which is the book on which the BLPD program is based. Seventy-nine percent of the control group principals reported reading the book by the third year of the study (up from 67% at baseline), compared with 83% of treatment principals. Thus, it appears, at least to some extent, that principals in the control group had some exposure to similar content as the treatment group principals, although the intensity of that exposure may have been limited compared with the intensity of the professional development provided directly by McREL to treatment principals.

In addition, in surveys that were administered in late winter of the third study year, we asked principals to indicate whether, in the past 12 months, they had participated in professional development on any of the following topics: reading, writing, mathematics, science, social studies, differentiated instruction, parent involvement, teacher collaboration, cooperative learning, or “other.” On average, principals in the control group reported that they had received professional development on 3.49 of these topics, whereas treatment principals received professional development on 3.29 of these topics. The difference between the two groups was not

statistically significant, suggesting the principals in the control group who did not attend the BLPD program, did receive alternative professional development opportunities. However, we do not know the quality or intensity of these experiences.

### *Impacts of the Program*

We were interested in assessing the impact of the program on principal leadership, the instructional climate of the school, principal efficacy, staff turnover, and student achievement. We begin here with a discussion of the proximal outcomes we measured with surveys: principal efficacy, principal leadership, and the school’s instructional climate (i.e., school climate and norms for teacher collaboration and differentiated instructional practice). Table 4 shows the results of the impact analyses on these outcome measures. The top panel of the table shows results based on principal surveys and the second panel displays results from teacher surveys. At the end of the third year of the study, principals in the treatment group reported feeling more efficacious than principals in the control group. The difference was statistically significant, with an effect size equal to .55.<sup>6</sup> Principals in the treatment group also rated themselves as more effective leaders than principals in the control group, although the difference was only marginally significant, with a  $p$  value equal to .10 and an effect size equal to .33. Similarly, principals in treatment schools reported more collaboration among staff (effect size = .40), a better school climate (effect size = .34), and stronger norms for differentiated instructional practice (effect size = .53) than principals in control schools. All differences were statistically significant at  $p < .10$ , and because effect sizes were greater than .25, can be considered substantively meaningful (What Works Clearinghouse, 2011).

However, there was no difference in the way teachers in treatment and control schools assessed the effectiveness of their principal’s leadership (second panel of Table 4). This suggests that while principals who attended the BLPD program felt they were exhibiting more effective leadership, the teachers did not experience it that way. Specifically, teachers in treatment schools

TABLE 4  
*Impact of BLPD on Principal Efficacy, Leadership, Instructional Climate, and Student Achievement*

Outcome variable	$\beta$	<i>SE</i>		<i>p</i> value	Effect size
Principal report ( <i>n</i> = 88)					
Principal efficacy	0.38**	0.15		.01	.55
Principal leadership	0.18	0.11		.10	.33
Collaboration	0.26*	0.14		.06	.40
School climate	0.21**	0.10		.03	.34
Collective differentiated instruction	0.44**	0.18		.01	.53
Teacher report ( <i>n</i> = 1,546 teacher in 91 schools)					
Principal leadership	0.05	0.12		.67	.03
Collaboration	0.07	0.06		.27	.06
School climate	0.02	0.05		.64	.02
Collective differentiated instruction	0.07	0.05		.17	.07

		ITT		TOT		<i>p</i> value	
Administrative data		$\beta$	<i>SE</i>	$\beta$	<i>SE</i>		
Teacher turnover ( <i>n</i> = 1,764)		−0.05*	0.03	−0.07*	0.04	.05	
Principal turnover ( <i>n</i> = 122)		−0.16*	0.09	−0.23*	0.13	.10	

		ITT		TOT		<i>p</i> value	ITT effect size
Student achievement	<i>n</i> of schools	$\beta$	<i>SE</i>	$\beta$	<i>SE</i>		
Math							
Grade 3	119	0.78	0.81	1.18	1.18	.34	.04
Grade 4	115	−0.01	1.03	−0.01	1.56	.99	.00
Grade 5	109	1.48	1.40	2.24	2.12	.29	.04
Reading							
Grade 3	119	−0.45	0.78	−0.68	1.19	.56	−.02
Grade 4	115	−0.27	1.01	−0.40	1.53	.79	−.01
Grade 5	109	0.81	1.25	1.23	1.89	.52	.02

*Note.* Covariates include % minority, % free and reduced lunch, school size, and average third-grade reading and math MEAP scores from 2008. All schools did not serve all grades in each year of the study, thus there are differences in the sample sizes across grades. TOT estimates adjusted by participation rate; effect size reflects both within- and between-level variance

(i.e.,  $\frac{\gamma_{01}}{\sqrt{\sigma^2 + \tau_{00}}}$ ). BLPD = Balanced Leadership® Professional Development; ITT = intent to treat; TOT = treatment on

the treated; MEAP = Michigan Educational Assessment Program.

\*Statistically significant at  $p \leq .10$  level. \*\*Statistically significant at  $p \leq .05$ .

did not report statistically significantly higher levels of principal leadership, teacher collaboration, school climate, or norms for differentiated instruction than teachers in control schools. Teachers were not asked to report on principals' sense of efficacy. In addition, although all of the effect size estimates of program impact on teachers' perceptions of principal leadership and

instructional climate were positive, they ranged from .02 to .07, and therefore cannot be judged as substantively meaningful (What Works Clearinghouse, 2011).

In contrast, the third panel of Table 4 shows a statistically significant impact of the BLPD program on both teacher and principal turnover. Specifically, there was a 16 percentage point

reduction in principal turnover and a 5 percentage point reduction in teacher turnover in the treatment schools. Estimates of the treatment on the treated are even higher—a 7 percentage point reduction for teachers and a 23 percentage point reduction for principals. The impact on principal turnover was striking—28 of the principals in the control condition had changed over the 3 years of the study as opposed to 14 of the treatment principals. Although we know from our qualitative interviews that some districts worked to keep principals at the same school so that they could continue to receive the BLPD program and that many principals commented on their high degree of satisfaction with the professional development, there were also cases in which districts worked to keep principals out of the training because they wanted to maximize the number of days they were in their schools. The estimates of treatment on the treated are even higher, suggesting that if all the principals had fully participated in the program, the reduction in principal turnover would have been 34 percentage points and the reduction in teacher turnover 9 percentage points.

Finally, the bottom panel of Table 4 shows the impact of the BLPD program on student achievement. There was no impact of the program on either reading or mathematics state achievement scores in third, fourth, or fifth grade. Moreover, none of the intent to treat effect size estimates exceeds an absolute value of .04. The limited impact on student achievement is not surprising given the results of the teacher survey, which indicated that teachers did not perceive improvements in principal leadership, school climate, or norms for collaboration and instructional practice in their schools as a result of the intervention.

In sum, the main statistically significant and substantively important impacts of the BLPD program were on the perceptions of treatment school principals regarding their sense of efficacy for instructional leadership, their leadership practices, the climate of their schools, and school-wide norms for teacher collaboration and instruction in their schools. In addition, the program had a causal impact on the rate of both principal and teacher turnover. However, teachers in treatment schools did not perceive improvements in leadership practice, school climate, or norms for their collaboration or use of practices

consistent with differentiated instruction and we found no impact of the program on student achievement. We turn next to a discussion of possible explanations for these mixed findings.

## Discussion

The McREL BLPD program is used widely to support the development of school leaders. The results from this study indicate that the program is generally implemented with a high degree of fidelity and that principals experience a high degree of satisfaction with the program. The principals who participated in the program reported feeling more efficacious, using more effective leadership practices and having a better instructional climate than principals in the control group. Prior research has shown that principals' perceptions are important predictors of effective leadership practices and academic climate (Leithwood & Jantzi, 2008; Urick & Bowers, 2011).

However, teacher reports indicate that principals' leadership practices and the instructional climate of the schools did not change as a result of the program. Furthermore, we found no impact of the program on student achievement. This is a disappointing result given the widespread use of the program.

There are a number of possible explanations for these findings. First, the BLPD program simply may not teach the necessary skills or behaviors to impact schools more positively than principals already are. It is quite possible that the meta-analysis, on which the BLPD framework is based, may itself have been flawed, because it included studies that suffered from omitted variable bias. This may have led to incorrect conclusions regarding the relationship between leadership and achievement. If so, then, the similarity of the content covered by the BLPD program and the content covered by most traditional educational administration programs and emphasized by the ISLLC standards would also call into question much of mainstream thinking in education administration. For a criticism of mainstream approaches to educational leader training, see Levine (2005).

Second, it may be that although principals found the professional development useful, the actual changes they made to their leadership

practice were small and had only a limited impact on the instructional climate of the school. For example, Barnes, Camburn, Sanders, and Sebastian (2010) have shown that principals who participate in professional development tend to “fine-tune” their practice rather than making large changes in their daily activities. Without substantial changes in practice, it is unlikely that we would observe changes in the way teachers experience the instructional climate of the school or in the overall level of student achievement.

Third, principals may have made substantial changes in their practice, but changes in leadership practice alone, without larger whole school reform or efforts that involve teachers and directly target the instructional climate of the school, may not result in impacts on student achievement. For example, Bryk et al. (2010) suggest that principal leadership is only one of five key essential supports required for school reform. Fourth, it is also possible that the time frame of the study did not permit us to observe changes to school climate and student achievement that occurred after the conclusion of the study. As Grissom et al. (2012) note, building an effective school environment may take several years to achieve.

There is also at least one design consideration that could have influenced the potential for the program to impact teachers’ perceptions and student achievement. The level of randomization was at the school and the 126 randomly assigned schools were nested in 74 distinct school districts. Thus, the modal district had only one participating school, thereby limiting the likelihood of district support for implementation, a concern voiced by at least some treatment school principals. A common reason for missing training sessions among treatment school principals was pressure from the superintendent to be in their schools more often or to be present for district meetings, which indicated instances of limited district support.

Finally, it may be that the BLPD program was effective, but that the leadership development that the control principals were receiving was equally effective and thus we find no difference between the treatment and control groups in terms of instructional climate or achievement. However, the differences in principals’ reports of their own practice call this interpretation into question.

The one area in which we did find an impact of the program was on principal and teacher turnover, with principals and teachers in treatment schools being significantly more likely to remain in the same school over the 3 years of the study than their control school counterparts. The intervention may have impacted turnover in a number of ways. For example, principals may have felt more satisfied in their jobs as a result of their participation in the program and thus may have been less inclined to leave. However, in exploratory analyses (not shown here), we find no evidence that increases in principal’s overall sense of efficacy or their own perceptions of their leadership practices mediated the relationship between BLPD participation and turnover.

However, districts may have simply been more likely to encourage their principals to stay in their positions or to refrain from moving them to another school to ensure that they were able to take advantage of the professional development opportunity being offered to them. There is some anecdotal evidence to suggest this may have been the case. Under this scenario, we would attribute the reduction in turnover not to participation in the BLPD program itself, but to the experiment or to the offer of free professional development.

There are several possible explanations for the reduction in teacher turnover as well. On one hand, teachers may have been less likely to leave a school if they felt that the principal was becoming more effective, but again, survey findings do not indicate that teachers perceived that their principals had changed substantially as a result of the training. On the other hand, consistent with previous literature (e.g., Bêteille et al., 2012), principal turnover may have mediated the impact of teacher turnover. In exploratory analyses not shown here, we do find some evidence that teachers were simply less likely to leave if their principal stayed. Although this finding does not bear directly on the BLPD program, it does suggest that any effort to retain principals and keep them at the same school could go a long way toward reducing teacher turnover, which has shown to be both costly and associated with lower levels of student achievement (National Commission on Teaching and America’s Future, 2007; Ronfeldt et al., 2013).



## Authors' Note

The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

## Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Roger Goddard was employed by McREL between September 1, 2012 and June 30, 2014. The study was designed and all data was collected prior to this appointment.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A080696 to the Texas A&M University.

## Notes

1. Our measure of socioeconomic status was the National School Lunch Program's Free or Reduced Price Lunch (FRPL) indicator found in the National Center for Education Statistics (NCES) Common Core of Data.

2. Designation 31—Town, Fringe: Territory inside an urban cluster that is less than or equal to 10 miles from an urbanized area. 32—Town, Distant: Territory inside an urban cluster that is more than 10 miles and less than or equal to 35 miles from an urbanized area. 33—Town, Remote: Territory inside an urban cluster that is more than 35 miles from an urbanized area. 41—Rural, Fringe: Census-defined rural territory that is less than or equal to 5 miles from an urbanized area, as well as rural territory that is less than or equal to 2.5 miles from an urban cluster. 42—Rural, Distant: Census-defined rural territory that is more than 5 miles but less than or equal to 25 miles from an urbanized area, as well as rural territory that is more than 2.5 miles but less than or equal to 10 miles from an urban cluster. 43—Rural, Remote: Census-defined rural territory that is more than 25 miles from an urbanized area and is also more than 10 miles from an urban cluster.

3. Three of the seven schools that closed or merged did so in the last year of the study, so we were able to observe turnover up until the time of the school closure and thus we included these principals and teachers in our turnover analyses.

4. Confirmatory factor analysis results, respectively, for teachers and principals, for all baseline surveys are as follows. Principal Leadership: comparative fit index

(CFI) = .99/.95 and standardized root mean square residual (SRMR) = .02/.05; Collaboration: CFI = .93/.99 and SRMR = .06/.05; Differentiated Instruction: CFI = .98/.95 and SRMR = .02/.04; School Climate: CFI = .93/.97 and SRMR = .06/.03; Principal Efficacy (from the principal survey only): CFI = .99 and SRMR = .03. Results were similar for final outcome surveys.

5. We could not control for baseline achievement at the student level because the state did not assess first or second graders in 2008. Thus, we do not know the baseline test scores of the second- or third-grade students in the fall of 2008 when they would have been in first or second grade. We did run models in which we controlled for student-level achievement for the fifth graders and found similar results to the ones shown here.

6. Effect sizes reflects both within- and between-level variance (i.e.,  $\frac{\gamma_{01}}{\sqrt{\sigma^2} + \sqrt{\tau_{00}}}$ ).

## References

- Allensworth, E. (2012). Want to improve teaching? Create collaborative, supportive schools. *American Education*, 36(3), 30–31. Retrieved from <http://files.eric.ed.gov/fulltext/EJ986682.pdf>
- Bandura, A. (1977). Self efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215. doi:10.1037/0033-295X.84.2.191
- Barnes, C. A., Camburn, E., Sanders, B. R., & Sebastian, J. (2010). Developing instructional leaders: Using mixed methods to explore the black box of planned change in principals' professional practice. *Educational Administration Quarterly*, 46, 241–279. doi:10.1177/1094670510361748
- Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, 41, 904–919. doi:10.1016/j.ssresearch.2012.03.003
- Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8, 225–246. doi:10.1177/0193841X8400800205
- Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, 48, 303–333. doi:10.3102/0002831210380788
- Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012, February). *Estimating the effect of leaders on public sector productivity: The case of school principals* (NBER Working Paper No. 17803). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w17803>
- Bryk, A. S., Bender Sebring, P., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago.

- Camburn, E. M., Goldring, E., May, H., Supovitz, J., Barnes, C., & Spillane, J. P. (2007). *Lessons learned from an experimental evaluation of a Principal Professional Development Program*. Retrieved from <http://www.cpre.org/lessons-learned-experimental-evaluation-principal-professional-development-program>
- Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 31, 92–109. doi:10.1016/j.econedurev.2011.09.001
- Council of Chief State School Officers. *Educational leadership policy standards: ISLLC 2008*. Retrieved from [http://www.ccsso.org/Documents/2008/Educational\\_Leadership\\_Policy\\_Standards\\_2008.pdf](http://www.ccsso.org/Documents/2008/Educational_Leadership_Policy_Standards_2008.pdf)
- DeAngelis, K. J., & White, B. R. (2011). *Principal turnover in Illinois public schools, 2001–2008* (Policy Research: IERC 2011-1). Edwardsville, IL: Illinois Education Research Council.
- Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37, 479–507. doi:10.3102/00028312037002479
- Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher*, 33(3), 3–13. doi:10.3102/0013189X033003003
- Goddard, R. D., Salloum, S. J., & Berebitsky, D. (2009). Trust as a mediator of the relationships between poverty, racial composition, and academic achievement evidence from Michigan's public elementary schools. *Educational Administration Quarterly*, 45, 292–311. doi:10.1177/0013161X08330503
- Goddard, R. D., Tschannen-Moran, M., & Hoy, W. K. (2001). A multilevel examination of the distribution and effects of teacher trust in students and parents in urban elementary schools. *The Elementary School Journal*, 102, 3–17. Retrieved from <http://www.jstor.org/stable/1002166>
- Goldring, E., Camburn, E., Huff, J., Sebastian, J., & May, H. (2007). *Effects of the National Institute for School Leadership: Early results from a randomized field trial*. Retrieved from [http://www.cpre.org/images/stories/cpre\\_pdfs/NISL%20Effects%20Early%20Results%20Final%20AERA%2007\\_final.pdf](http://www.cpre.org/images/stories/cpre_pdfs/NISL%20Effects%20Early%20Results%20Final%20AERA%2007_final.pdf)
- Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, 113, 2552–2585.
- Grissom, J. A., Kalogrides, D., & Loeb, S. (2012, November). *Using student test scores to measure principal performance* (NBER Working Paper No. 18568). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w18568>
- Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33, 235–261. doi:10.3102/0162373711398128
- Leithwood, K., & Jantzi, D. (2008). Linking leadership to student learning: The contributions of leader efficacy. *Educational Administration Quarterly*, 44, 496–528. doi:10.1177/0013161X08321501
- Leithwood, K., & Seashore Louis, K. (2012). *Linking leadership to student learning*. San Francisco, CA: Jossey-Bass.
- Levine, A. (2005). *Educating school leaders*. New York, NY: Education Schools Project.
- Marzano, R. J., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, 36, 60–72. doi:10.1016/j.econedurev.2013.05.004
- National Commission on Teaching and America's Future. (2007). *The high cost of teacher turnover* (Policy brief). Available from <http://www.nctaf.org>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.
- Robinson, V., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, 44, 635–674. doi:10.1177/0013161X08321509
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, 50, 4–36. doi:10.3102/0002831212463813
- Urlick, A., & Bowers, A. J. (2011). What influences principals' perceptions of academic climate? A Nationally Representative Study of the Direct Effects of Perception on Climate. *Leadership and Policy in Schools*, 10, 322–348. doi:10.1080/15700763.2011.577925
- Wallace Foundation. (2013). *The school principal as leader: Guiding schools to better teaching and learning*. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/effective-principal-leadership/Documents/The-School-Principal-as-Leader-Guiding-Schools-to-Better-Teaching-and-Learning-2nd-Ed.pdf>
- Waters, J. T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.

Weinstein, M., Schwartz, A. E., Jacobowitz, R., Ely, T., & Landon, K. (2009). *New schools, new leaders: A study of principal turnover and academic achievement at new high schools in New York City* (NYU Wagner Research Paper No. 2011-09). Retrieved from <http://ssrn.com/abstract=1875901>

What Works Clearinghouse. (2011). *What Works Clearinghouse procedures and standards handbook* (v. 2.1). Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v2\\_1\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf)

### Authors

ROBIN JACOB is a research assistant professor at the Institute for Social Research and the School of Education at the University of Michigan. Her research focuses on evaluations of education interventions and evaluation methods. She has a special interest in how policies and programs can affect instructional quality and outcomes in elementary schools.

ROGER GODDARD is the Novice Fawcett Chair in educational administration at The Ohio State University. His research interests include the social psychology of school leadership and organization, educational measurement, and advanced statistical analysis, with a particular focus on social cognitive theory and the study of efficacy beliefs.

MINJUNG KIM is a postdoctoral research fellow in the Department of Psychology at the University of South Carolina. She received her PhD from Texas A&M University in the educational psychology (Research, Measurement, and Statistics program). Her research interest includes longitudinal data analysis utilizing multilevel modeling and structural equation modeling. She is also interested in assessing individual differences using regression mixture models.

ROBERT MILLER most recently served as an associate research scientist at Texas A&M University—College Station. He has extensive experience in program evaluation and conducting analysis of large-scale education databases. His work focuses on education policy, organization theory, and the analysis of school effectiveness.

YVONNE GODDARD is a visiting associate professor at The Ohio State University. Her research interests include effective literacy instruction for students with disabilities and who are at-risk as well as teacher collaboration as it relates to efficacy beliefs, instructional practices, and student outcomes.

Manuscript received October 14, 2013

Revision received April 30, 2014

Accepted July 14, 2014