Can Professional Development Improve School Leadership? Results From a Randomized Control Trial Assessing the Impact of McREL's Balanced Leadership Program on Principals in Rural Michigan Schools

Robert James Miller[1], Roger D. Goddard[2], Minjung Kim[3], Robin Jacob[4], Yvonne Goddard[2], and Patricia Schroeder[5]

Abstract

Purpose: This multiyear experimental study was designed to examine (1) the causal impact of McREL International's Balanced Leadership® Professional Development (BLPD) program on school principals' learning, beliefs, and behaviors and (2) whether there were differences in the types of outcomes the professional development influenced. Outcomes included principals' reported sense of efficacy, perceptions of school climate, and leadership behaviors. **Research Methods:** Approximately 100 school principals in rural Michigan were randomly assigned to either a treatment

[1]Edgewood Independent School District, San Antonio, TX, USA
[2]The Ohio State University, Columbus, OH, USA
[3]University of Alabama, Tuscaloosa, AL, USA
[4]University of Michigan, Ann Arbor, MI, USA
[5]Independent Researcher, TX, USA

Corresponding Author:
Roger D. Goddard, 2561 Sherwin Rd, Upper Arlington, OH 43221, USA.
Email: goddard.9@osu.edu

group that was offered 2 years of BLPD training or a control group that conducted business as usual. We employed principal survey data from the first and third years of the School Leadership Improvement Study. The treatment effect on each outcome was assessed through regression analysis, which controlled for baseline scores and school demographics. **Findings:** Results from the School Leadership Improvement Study show that BLPD participants reported substantively significant growth on the majority of the outcomes targeted by the program. Interestingly, treatment principals were more likely to report growth on broad, school-level outcomes than in areas that involved them working directly with teachers. **Implications:** This evaluation demonstrates that the McREL Balanced Leadership program caused gains in the majority of knowledge, belief, and behavior outcomes with the largest impacts on principals' sense of efficacy for instructional improvement, reported ability to bring about change, and strength of norms for teachers' instructional practice. The authors explore possibilities for why broad, school-level outcomes were more likely to be affected than areas that involve principals more directly in teachers' work.

Introduction

The role of principals in leading school-based instructional improvement efforts is receiving increased attention as accountability standards and changing demographics continue to challenge American schools (Hallinger, 2003, 2005). Supporting calls for improved school leadership is research evidence demonstrating that leadership was a primary driver of school improvement in Chicago (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). In addition, meta-analytic studies have shown that school leadership is statistically linked to student outcomes (Robinson, Lloyd, & Rowe, 2008; Waters, Marzano, & McNulty, 2003). Other research suggests that measures of principal quality can effectively predict student achievement (Owings, Kaplan, & Nunnery, 2005). Complimentary work by Darling-Hammond, Meyerson, LaPointe, and Orr (2009) suggests that schools simultaneously require effective leaders for their success but that school improvement efforts suffer "in part due to a lack of supports for developing such leadership" (p. VII). Where effective leaders exist, other work indicates their distribution across school contexts is uneven. Specifically, research by Branch, Hanushek, and Rivkin

(2012), released by the National Bureau of Economic Research, indicates that there is wide variation in principal quality among schools serving poor children. These works suggest that while principal quality is important, we know too little about how it may be developed and that principal quality for the largely poor rural schools studied here likely varies widely. For these reasons, we conducted a randomized control trial to ask whether a research-based training program may help principals in poor rural schools grow in productive ways.

The growth of research evidence suggesting that principals matter to school improvement and student learning supports consistent calls for high-quality school leadership including, for example, the U.S. Department of Education's (2010) statement that every school needs a "great" principal as outlined in its *Blueprint for Reform*. At the same time, revisions to standards for educational leader licensure[1] not only continue to underscore the relationship between school leadership and student success but also make a strong case that instructional leadership is one of the most important roles a principal can fulfill to support national goals for educational equity and excellence (Council of Chief State School Officers, 2014).

Unfortunately, however, calls for exceptional leadership in U.S. schools are far more common than research-based knowledge about how to develop such leaders to fill these needs. For example, Levine (2005) posits that traditional university-based doctoral programs have an uneven and largely negligible record of demonstrating effectiveness in preparing principals to lead schools. Furthermore, principal preparation programs and in-service professional development training opportunities often lack consensus on the range of skills and knowledge principals need to be successful leaders (Jackson & Kelley, 2002; Knapp, Copland, & Talbert, 2003). As Darling-Hammond et al. (2009) observe, although compelling evidence underscores the importance of leadership to instructional improvement, school climate, and school effectiveness, effective leadership training programs are not widely available and few training programs have rigorous research evidence of their effectiveness.

There is some emerging work in this area, however. For example, Darling-Hammond et al. (2009) examined commonalities among several diverse programs for preservice and inservice leadership professional development. Based on their extensive reviews of these programs, Darling-Hammond et al. concluded that there are shared aspects of exemplary principal training programs. Across programs, several common themes emerged, including standards-aligned curriculum, integrating theory and practice that involves field time, knowledgeable faculty and practitioners, and social and professional supports. Also, Hamilton, Engberg, Steiner, Nelson, and Yuan (2012) conducted a mixed methods study of the Pittsburgh

Principal Incentive Program to examine program implementation, how principals and other school staff responded to the program, and other outcomes associated with program implementation including student achievement. Although the authors could not offer definitive evidence of the effects of this program, they were able to provide several recommendations regarding principal evaluation, support, and compensation policies. These included obtaining regular input from a range of stakeholders, calibration training to assure evaluator consistency, monitoring ongoing student achievement gaps, aligning reforms with teaching and learning expectations, helping principals find the time to engage in expected practices, and examining connections between principal mobility and school effectiveness. Previously, Leithwood, Riedlinger, Bauer, and Jantzi (2003) found that a leadership fellows program, offered by a not-for-profit center, emphasizing instructional improvement and transformational leadership was associated with modest increases in student performance and teacher perceptions of principal effectiveness, and that these improvements were driven largely by leaders' ability and training to influence school conditions. Furthermore, Grissom and Harrington (2010) used teacher-level survey responses from the School and Staffing Survey related to school management and found that principals who engaged in formal mentoring and coaching activities were assessed as being effective, whereas a negative, causal relationship existed between teacher reports and principal effectiveness when principals took university coursework as professional development. Furthermore, they found that 33% of rural principals reported participating in mentoring or coaching as opposed to those in urban (55%) or suburban (39%) settings. The authors postulate that these discrepancies exist because of the relatively smaller number of schools, and therefore principals, in rural versus urban settings who can serve as mentors or coaches. Across these three studies, several themes emerged, most of which do not overlap. On closer examination, however, principals' collaborative work (Hamilton et al., 2012), social and professional supports (Darling-Hammond et al., 2009), and mentoring/coaching (Grissom & Harrington, 2010) can be construed as related constructs. Thus, across several studies, much research remains to determine which aspects of principal professional development are supported empirically.

McREL International (McREL) developed the Balanced Leadership® Professional Development (BLPD) program based on research describing the knowledge, beliefs, and behaviors of school principals that have been shown by extant research to be related to student achievement. The research we report here was designed to advance knowledge by employing an experimental design to investigate how principals' leadership capacity does and does

not develop through professional development while on the job. More specifically, we examined patterns across the impacts of McREL's research-based principal professional development program based on principal reports in the domains of principal efficacy beliefs, principal leadership, school climate, norms for teacher collaboration, and the use of effective instructional strategies. In addition to looking across domains, we examined differences in the magnitude of causal impacts within the domains of principal leadership and school climate.

In sum, we were interested in learning what types of growth did and did not occur as a result of principals participating in McREL's intensive, cohort-based professional development program over 2 years. Our macro questions (i.e., across domains) included whether principals' perceptions of their own leadership were more likely to grow than their beliefs about the climate of their schools, the collective practices of their teachers, or their own sense of efficacy for instructional leadership. Our micro questions (i.e., within domains) addressed whether there were differences in principals' perceived growth within the broad domains of principal leadership and school climate.[2]

Importantly, the purpose of our research was neither to defend nor discredit the content of the McREL BLPD program. Rather, we simply observe in our literature review that the program's elements have research-based support and that its delivery and content were sufficiently well-organized and developed to warrant a federal investment into investigating its probable causal impacts. Thus, the work of McREL to develop a comprehensive research-based training program for school leaders provides a unique opportunity to evaluate experimentally the efficacy of a prominent professional development program reasonably designed to achieve its goals. We turn next to a review of the conceptual and empirical foundations of the BLPD program and the School Leadership Improvement Study (SLIS) that was designed to evaluate its effects.

Balanced Leadership Background and Theoretical Framework

The leadership behaviors and practices emphasized by McREL were developed based on a meta-analysis of leadership studies conducted over a span of more than three decades (Waters et al., 2003). Through their meta-analytic synthesis of this research, McREL identified 21 school-level leadership responsibilities previously linked with increased levels of student achievement in districts and schools. The leadership responsibilities identified by McREL are listed in Appendix B.

The 21 responsibilities are drawn collectively across 69 separate studies meeting a priori quality criteria, with each study representing a range or subset of the designated responsibilities rather than a complete model of leadership as presented by the McREL BLPD program. Thus, although the BLPD training is based on a heuristic framework, Balanced Leadership® itself is not so much a grand theory of educational administration as it is a compilation and synthesis of *what works* in education leadership research. Importantly, the leadership responsibilities enumerated by McREL have empirical support from the extant principal leadership literature. Specifically, Waters et al. (2003) found across the meta-analyzed studies an average statistically significant, positive relationship between school-level leadership and student achievement of 0.25 *SD*, which represents approximately 10-percentile points on norm-referenced tests of academic achievement. Although Waters et al. included some unpublished works among the reviewed studies, this effect size finding was supported by a more recent peer-reviewed meta-analysis of leadership effects on student achievement reported by Robinson et al. (2008). Common across the meta-analyses of Waters et al. and Robinson et al. is empirical support for dimensions of principal leadership and behavior that are both potentially malleable through professional development learning and positively linked to increased student outcomes.

Principal leadership is heavily woven into and represented by virtually all of the 21 responsibilities. In addition to the leadership responsibilities that are expected to guide principals' day-to-day practice and behavior, the McREL BLPD program emphasizes a number of broad theoretical concepts and processes that encompass the responsibilities. Specifically, via the 21 leadership responsibilities, the BLPD framework focuses on improving school-level leadership related to the following three constructs: *purposeful community, magnitude of change*, and *focus of change*.

Purposeful community refers to collective efficacy (Goddard, 2001; Goddard, Hoy, & Woolfolk Hoy, 2004; Waters et al., 2003) among the principal and staff as they work as a community to build the capacity to accomplish purposes and produce outcomes through consensus and agreed on processes. Furthermore, in a purposeful community the principal is encouraged to use social persuasion and shared leadership to promote teacher collaboration concentrated on instructional improvement.

The concept of change is also central to the BLPD program. By emphasizing *magnitude of change*, the program stresses principals' understanding of the degree of disequilibrium the changes they pursue in their schools may cause. Roughly paralleling what Argyris and Schön (1974, 1978) termed *single-loop and double-loop learning*, McREL emphasizes the difference between what they describe as first- and second-order change. In both cases the distinction lies in

whether the change at hand requires deviance from current organizational norms, practices, and/or policies (second-order change) or whether the change can be accomplished without challenging prevailing paradigms (first-order change).

Regardless of the magnitude of a given change, the BLPD program strongly emphasizes principals' understanding of the *focus of change* in their schools. Such an emphasis is consistent with other recent research supporting the crucial role principals play in school change (e.g., Fullan, 2011) and the importance of principals' support for change to teachers' work (Berebitsky, Goddard, & Carlisle, 2014). Principals are expected to act as change agents by systematically seeking new and better ways of effective schooling. Thus, as is the case with research literature supporting the 21 leadership responsibilities, the broader McREL BLPD theoretical concepts also reflect research-based support linked to increased student achievement.

Notably, many of the elements and outcomes of the BLPD program draw considerable support from research published *since* McREL completed the meta-analysis on which their program is based. For example, Goddard, Hoy, and Woolfolk Hoy (2004) and Goddard, Goddard, Kim, and Miller (2015) have demonstrated that collective efficacy, the beliefs of teachers in a school that they will be able to organize and execute the courses of action required to successfully educate the students in their charge, is a predictor of student achievement across a number of school contexts. Social trust, too, which is reflective of a purposeful community and refers to the degree to which relationships are characterized by honesty, openness, benevolence, reliability, and competence, is supported by recent research showing its impact on student achievement (Bryk et al., 2010; Bryk & Schneider, 2002; Goddard, Salloum, & Berebitsky, 2009). In addition, Lee and Smith (1996) and Goddard and LoGerfo (2007) have shown that collective responsibility, which embraces the sense of resolve found in a purposeful community, is related to student achievement. Collective responsibility refers to the degree to which teachers believe their work is the cause of the outcomes students' experience. And, while Hausman and Goldring (2001) argued that meaningful change requires teacher input (one of the 21 responsibilities), more recent evidence shows that the extent to which teachers collaborate around instructional improvement is directly related to student academic outcomes (Goddard, Goddard, & Tschannen-Moran, 2007; Louis, Dretzke, & Wahlstrom, 2009). Collaboration has also been shown by Goddard, Goddard, Kim et al. (2015) to mediate the relationship between leadership and achievement in schools. Finally, recent research on differentiated instruction (a probable outcome of monitoring, evaluation, involvement in instruction, and support for innovation) indicates it is positively and significantly related to student achievement (Goddard, Goddard, & Kim, 2015).

In sum, there is a research base for the leadership responsibilities and concepts included in the McREL BLPD program. Moreover, research conducted since McREL's 2003 meta-analysis provides continued support for the program's core elements and the statistical link between school leadership and student achievement (Robinson et al., 2008). To date only Jacob et al. (2015) have examined the impacts of McREL's unique professional development program but they did not provide a detailed examination of outcomes across and within the domains examined here. To set the stage for a discussion of our evaluation framework, we first briefly review the way the BLPD program was delivered to participants in our study.

Balanced Leadership® Program Delivery

The context of this study was "economically-disadvantaged" rural schools in Michigan. To assess the causal impact of the program, half of a group of school leaders who agreed to participate was randomly assigned to receive BLPD program for approximately 2 years (the treatment group) while the other half was assigned to conduct business as usual with no intervention (the control group). As described more fully in the Method section, there were almost 100 schools across both groups with principals who participated. Because the treatment group participants were spread over a large region of northern Michigan that would make their participation in the same McREL-led professional development sessions unmanageable, the participants strategically chose two locations that allowed all participants to attend cohort sessions with a reasonable commuting time. In this way, treatment participants were organized into two cohort groups of 20 to 30 principals with roughly equal control school groups in the same geographic area. Each treatment group was offered BLPD training over the course of 2 calendar years through an intensive series of 10, 2-day professional development sessions to help practitioners learn and apply the content of the McREL BLPD program. The content was delivered by McREL's trained professional development facilitators with deep knowledge of the program content and extensive prior experience with its delivery. Importantly, the content, length, and manner of delivery for McREL's BLPD program were unaltered for purposes of this study.

McREL's program delivery model is designed to allow participants the opportunity to implement what they learned in training at their respective school sites between sessions. The delivery design also allows participants to share their reflections about the training and about problems of practice encountered at their schools (network reflection) with other session cohort members and the program providers. The overall intent of the training is to help principals gain not only declarative and procedural knowledge of

McREL's 21 responsibilities, but also to permit experiential and contextual opportunities to apply what they have learned.

In sum, the McREL Balanced Leadership® program has a research base and a well-developed program delivery protocol. Prior to this evaluation, however, McREL's BLPD had not undergone rigorous external evaluation to examine its effects on principal behavior and school improvement. Therefore, we designed this study to assess the causal impact of McREL's BLPD program on principals' professional development learning. We turn next to the framework employed to guide our evaluation of McREL's school leadership development approach.

Evaluation Framework

To examine the intervention fidelity and efficacy of the McREL Balanced Leadership® program, we refer to Kirkpatrick's taxonomy for evaluating training programs (Kirkpatrick, 1959, 1994). Kirkpatrick identifies four levels of criteria for evaluating training programs: (1) reaction, (2) learning, (3) behavior, and (4) results. *Reaction* refers to participants' program receptivity in terms of affective reactions (i.e., satisfaction) and utility judgments such as perceived effectiveness or merit of a training program (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). *Learning* measures the attainment of knowledge, principles, strategies, and techniques specified by training objectives. *Behavior* refers to positive changes in job performance directly related to training program objectives. Finally, *results* represent the desired organizational outcomes based on the routine enactment of training skills and objectives specified by the training program.

Program receptivity (*reaction*) to BLPD was analyzed in companion work that confirms principal satisfaction with McREL's program delivery (Schroeder et al., 2012). At near full consensus, self-reports by the principals indicate that the program contributed to their understanding of school leadership, and that the principals intended to use their newly acquired knowledge in their practice. We address Kirkpatrick's *learning* and *behavior* criterion in this article. *Results*, an examination of BLPD program effects on the instructional climate of the school from the perspective of the teachers, staff turnover, and student achievement, is examined in a separate article (Jacob, Goddard, Kim, Miller, & Goddard, 2015).

Here, we (the independent, university-based research team) examine the extent to which treatment group principals differ from control group principals on their reports of knowledge, beliefs, and behaviors postulated as outcomes of the McREL Balanced Leadership® program. This aspect of our work focuses on Kirkpatrick's (1959, 1994) *learning* criteria. To determine

learning, we attempted to measure the extent to which participants increased knowledge, changed beliefs, and/or increased effective school leadership behavior as a result of participating in the BLPD program. An example of this is principals' reports regarding their knowledge of curriculum, instruction, and assessment. Next, to address Kirkpatrick's (1959, 1994) *behavior* criteria, we examined school leaders' self-reported use of BLPD-related tools, practices, and behaviors, as well as their assessments of school climate and norms for teacher practice that may have been influenced by their school leadership as a result of program enactment. Examples of this include principals' monitoring of classroom instruction and outreach to school stakeholders.

It is also important to determine whether principals maintained and transferred their learning beyond the final McREL training session and generalized or applied their new knowledge and skills at their schools. Maintenance and generality of learning are important constructs that measure the effects of training over time (maintenance) and into new settings and/or situations (generality). For example, Sheeler, Bruno, Grubb, and Seavey (2009) found that teachers who were trained using immediate feedback and prompting when planning how to implement newly learned skills to their classroom settings effectively maintained these skills over time and generalized the skills to their classrooms as compared with teachers who did not receive such explicit training. Similarly, McREL's Balanced Leadership® program involves immediate feedback as participants plan how to implement new knowledge in their specific settings. Thus, we expected maintenance and generality and collected outcome data 1 to 2 months after the final training session on site in the participant's schools.

In sum, our general hypothesis was that the McREL program would result in growth in principals' reported knowledge, beliefs, and behaviors. The domains we examined as outcomes included principal leadership, school climate, principal efficacy beliefs, and school norms for teacher collaboration and use of differentiated instruction. Notably, these outcomes include measures of many of the 21 responsibilities contained in the program as well as measures of some of the program's probable outcomes. Because of the limits imposed by measurement science and survey length considerations, we could not measure all aspects of the program's four domains (i.e., purposeful community, magnitude of change, focus of change, and the 21 responsibilities). Given this, we relied on extant research and theory to guide our deployment of surveys likely to yield valid and reliable measures of plausible program outcomes. In addition, we planned no tests of relationships among the outcomes; rather, we simply examined the degree to which treatment and control school principals differ on these outcomes controlling for their

baseline (pretreatment) status on each indicator. A logic model describing the program and the mediating outcomes we assessed appears in Appendix A.

In addition to hypothesizing that treatment site school principals would on average report statistically higher than control site principals across BLPD-related outcomes, we were particularly interested in whether there were differences across the outcome domains in the effect of the BLPD program. Because there is little research in this area, these questions were exploratory and therefore we did not offer directional hypotheses regarding whether principals should experience greater growth in perceptions of their own leadership behavior versus, for example, their sense of efficacy for instructional leadership or their perceptions of school climate. An added benefit of having multiple measures of school climate and leadership was that we were able to look within these domains, again in an exploratory fashion intended to afford the opportunity to develop knowledge regarding whether themes explaining differential gains emerged.

Method

This section describes the study sample, participants, measures, and primary analytic methods employed to test our main hypothesis. It should be noted that the SLIS is an independent, university-based, federally funded Institute of Education Sciences research project designed to study the treatment fidelity and effectiveness of the McREL Balanced Leadership® program in producing principal leadership behavior changes, school climate changes, and ultimately student learning. McREL solely developed the program and was responsible for the delivery of training while the SLIS research team worked independently to evaluate the program.[3]

Sample and Participants

The sample of schools for this study was drawn from the population of Michigan's rural elementary schools, which on average serve high proportions of students receiving a free or reduced price lunch and have large percentages of students performing below state expectations on the Michigan Educational Assessment Program. To be included in our sample, schools were required to be located in a district with a Census Bureau Locale code ranging from small town to rural. Small cities and larger metropolitan areas thus fell outside the Census Bureau codes used to define the population for this study. Among qualifying schools located within driving distance of two predetermined offsite training locations, we recruited a total of 126 qualifying schools[4] that initially agreed to participate in the study. At baseline collection

in late fall of 2008, a total of 95 schools participated in the study and an analysis of attritor schools presented in the results section showed no difference in school demographic composition and average achievement from the original frame. Among principals located in the participating schools, the group averaged 9.75 years of service as principals and averaged 5.91 years of service at the present school site. The majority of principals (94%) previously served as teachers during their education career with about 60% indicating teaching experience at the elementary school level.

Training Timeline and Data Collection

Balanced Leadership training was offered across 2 years beginning in January of 2009 and ending after a series of 10 two-day sessions in October of 2010. Baseline data collection was completed in November and December of 2008 before BLPD training began. Final outcome data were collected in late fall and winter of the 2010-2011 school year after the treatment principals completed advanced training sessions on the extension and refinement of the 21 BLPD leadership responsibilities emphasizing school focus, change, and purposeful community as a system of school improvement.

Measures

The evaluation team employed principals' survey data from the first and third years of the SLIS. The principal survey was designed to measure the proximal outcomes targeted by the BLPD program as well as a number of other outcomes that, while not identified by the program developer as among the 21 leadership responsibilities, were nevertheless probable outcomes of successful enactment. Thus, the survey items presented on the questionnaire represent a comprehensive, but not exhaustive, attempt to measure likely outcomes of the BLPD program. And, while a few BLPD responsibilities are not directly represented on the principal survey (affirmation, communication, contingent rewards, optimize, order, resources, and relationships), other likely BLPD outcomes with a strong research base such as collective efficacy and teacher collaboration for instructional improvement are included. We took this approach to meet our goal of assessing the impact of the McREL BLPD training program on a range of probable outcomes that research has identified as important to school improvement.

Our pre- and posttraining principal survey included 27 measures of principal knowledge, beliefs, and behavior, including McREL-defined responsibilities and measures of school climate and norms for teacher practice. Specifically, the SLIS principal survey scales based on anticipated

outcomes of the Balanced Leadership® training included discipline; collective focus on student learning; principal's involvement in curriculum, instruction, and assessment; outreach; principal as change agent; flexibility; intellectual stimulation; principal's knowledge of curriculum, instruction, and assessment; school norms for differentiated instruction; culture; teacher collaboration; parent and community involvement; principal's visibility; principal support for change; collective efficacy; network reflection; social persuasion; principal leadership; collective responsibility; principal trust in teachers; principal efficacy; situational awareness; principal leads change; guaranteed and viable curriculum; monitor and evaluate; ideals and beliefs; and purposeful community. These measures span a variety of related but distinct conceptualizations of school climate and principal leadership. For example, while collective *efficacy* refers to teachers' beliefs in their conjoint capability to successfully educate the students in their charge, collective *responsibility* refers to the degree to which teachers view themselves as responsible for whether or not students learn. Similarly, in the area of principal leadership, change *agent* refers to principals managing the conditions necessary to work through change, including challenges to the status quo, whereas *leading* change specifically asks whether the principal has led any major change efforts in the past 12 months. The reader will find sample items for all 27 subscales in the presentation of the results. To assess the psychometric properties of these measures, we employed exploratory and confirmatory factor analysis to confirm the unidimensionality of the scales in both study years and Chronbach's alpha to assess the reliability of scores on these scales. With a total of 27 subscales, the analytic work presented here is a fairly exhaustive attempt to represent the leadership responsibilities and concepts comprising probable outcomes of the McREL program.

Data Analysis

To evaluate the effects of the multi-session BLPD program, we conducted two analyses. First, changes between pre- and postsurvey gain score means were analyzed for all outcomes, including McREL-defined responsibilities and related concepts measured in the SLIS principal survey. Second, the treatment effect on each outcome was assessed through regression analysis, which controlled for each principal's baseline score on each outcome, along with school-level demographic characteristics. Specifically, we estimated the following ordinary least squares model:

$$Y_j = \beta_0 + \beta_1 T_j + \sum_{2}^{x} \beta_x COV_{xj} + \varepsilon_j,$$

where Y is a measure of principal learning or behavior; T is an indicator variable equal to 1 if the principal is in a treatment school and 0 otherwise; β_1 the average impact of the program for the schools in our sample; COV are a list of school-level covariates including a baseline measure; and ε_j is the residual error term.

The demographic control variables included total enrollment in school, percent minority enrollment, percent students receiving free and reduced lunch, and 2008 state assessment reading and math scaled scores for third grade.[5] Thus, the coefficient on T will indicate the impact of the BLPD program on each of the outcomes assessed net of the effect of school demographics and prior achievement status. Based on the regression analyses, we also produced effect size estimates to compare differences in outcomes between the treatment and control school principals in response to the McREL BLPD training. Hedges' g was used as the method to estimate effect sizes as this procedure is recommended for small sample sizes (Grissom & Kim, 2005).

Finally, it is important to observe that offering the Balanced Leadership® program to principals in a particular group of schools does not guarantee that all principals will fully participate in the program. Principals may leave study schools at some point during the intervention or may simply decide not to attend some or all of the Balanced Leadership® training sessions. By estimating impacts on all principals for whom we obtained survey data, regardless of whether or not the principals actually participated in the Balanced Leadership® program, we were able to obtain estimates of the intent to treat (ITT) or the impact of *offering* the program to principals, represented by the coefficient β_1 above. Because principal turnover and noncompliance (e.g., not attending some or all of the training sessions) are very real and common occurrences in schools, the ITT estimate answers a policy-relevant question, namely, the impact a district implementing the Balanced Leadership® program in their schools can expect to obtain given a certain amount of principal turnover or noncompliance. However, these estimates cannot tell us what the impact of the program would be if all principals had participated fully in the Balanced Leadership® Training, which is also an important policy-relevant question. We therefore also estimated what is often referred to as the effect of the treatment on the treated (TOT), using a correction suggested by Bloom (1984), in which the ITT estimate is scaled up by the proportion of principals who fully participated in the program. Such analyses provide better estimates of the effect of the treatment in schools where the principal fully participated in the Balanced Leadership® program than controlling for attendance, which introduces bias because attendance cannot be assumed to be exogenous to the intervention.

Table 1. Descriptive Statistics of the Schools at Baseline (*N* = 126).

	Participants (*N* = 95)				Nonparticipants (*N* = 31)				*t* Value	*p* Value
	Mean	*SD*	Min.	Max.	Mean	*SD*	Min.	Max.		
MEAP Reading[a]	332.79	6.78	307.50	346.65	331.44	5.96	319.80	344.50	.99	.327
MEAP Math[b]	330.62	8.18	308.00	350.43	330.39	9.97	313.38	360.00	.13	.898
School size[c]	301.82	145.57	36.00	1063.00	302.00	168.67	30.00	704.00	−.01	.995
% Minority	0.11	0.15	0.00	0.74	0.10	0.19	0.00	0.95	−.89	.376
% Free/ reduced price lunch	0.51	0.15	0.14	0.83	0.53	0.13	0.26	0.77	−.66	.512

Note. MEAP = Michigan Educational Assessment Program.
[a]2008 MEAP Reading scaled scores for third grade. [b]2008 MEAP Math scaled scores for third grade. [c]The number of total enrollment in school.

Results

In this section, we provide sample descriptive statistics, a summary of the psychometric properties of our measures, findings relative to differential attrition, missing data and imputation procedures, and the results of our program impact analyses.

Descriptives

Descriptive statistics for these schools and correlations among demographic variables describing the schools in our sample are reported in Tables 1 and 2, respectively. Importantly, Table 1 shows that after random assignment the average composition of treatment schools was not statistically different from that of control schools in terms of prior achievement, size, ethnicity, or socioeconomic status. In addition, just over half the students in each group received a subsidized lunch. As Table 2 indicates, the variable most strongly associated with achievement was the percentage of students receiving a free and reduced price lunch.

Psychometric Analysis

The psychometric properties of the 27 outcomes measured by the SLIS principal survey were established through exploratory and confirmatory factor analysis as well as reliability measures. These analyses confirmed the unidimensionality of all 27 scales across study years and the results. This is not surprising given that the majority of the survey measures have been published

Table 2. Correlations Among the Variables for School Characteristics at Baseline (*N* = 95).

	MEAP Reading[a]	MEAP Math[b]	Treatment	School Size[c]	% Minority
MEAP Reading[a]	1.000	—	—	—	—
MEAP Math[b]	.678**	1.000	—	—	—
Treatment[c]	−.010	.015	1.000	—	—
School size[d]	.154	.148	−.023	1.000	—
% Minority	−.094	−.084	−.078	−.097	1.000
% Free/reduced price lunch	−.270**	−.337**	−.029	−.349**	.126**

Note. MEAP = Michigan Educational Assessment Program.
[a]2008 MEAP Reading scaled scores for third grade. [b]2008 MEAP Math scaled scores for third grade. [c]Treatment schools are coded as 1 while the control schools are coded as 0. [d]The number of total enrollment in school.
**Statistically significant correlation at *p* < .01. *Statistically significant correlation at *p* < .05.

previously or developed by McREL over time and prior to the beginning of this evaluation. To illustrate, there were almost no items for subscales with factor loadings below .70 and none below .65, and most loadings were in the range of .75 to .95. The corresponding internal consistency reliabilities (Cronbach's alpha) were moderate to high, ranging from a minimum of .64 to a maximum of .94, with a mean reliability of .84. Because these results and those of related confirmatory factor analysis are summarized in a companion article (Miller et al., 2013), they are not presented in detail here.

Attrition, Compliance, and Missing Data Analyses

Because the ecological reality of a long-term training program in schools is that personnel changes do occur, it is policy-relevant to know the outcomes of programs not just in schools without turnover but also in all schools. To produce accurate estimates of principal growth on the McREL leadership responsibilities and related program outcomes, we applied multiple analytic strategies to examine differential attrition, principal turnover, and missing data. We turn to these analyses next.

We begin by exploring attrition. Between random assignment of our initial sample of 126 schools and the first wave of data collection, we lost a total of 31 schools, all of which either actively refused to participate once the results of random assignment were known or who failed to complete baseline surveys. Of these 31 schools, 19 were assigned to the treatment

condition and 12 to control. This resulted in a total sample of 95 schools (43 treatment and 52 control). Overall attrition was equal to 25%. We lost slightly more treatment schools (31%) than control schools (19%), for a differential response rate of approximately 12 percentage points.[6] A comparison of the 95 participating schools to the 31 that opted out indicated no statistically significant differences on any demographic or prior achievement variables, suggesting that attrition was not systematically related to school context or achievement status.

Among these 95 schools, 29 experienced principal turnover between the beginning and end of the project. In addition, of the 95 participating schools, 88 provided both pre- and postsurvey responses. Among the 88 schools reporting data for both data collection waves, almost three-quarters (62) experienced no principal turnover. We performed a series of statistical tests to determine if there were any significant differences among model variables in the missingness of data for all possible comparisons (e.g., the 95 original schools compared with the 88 for which data were present at both time points, the 95 vs. the 62). Our results showed that there were no statistically significant differences between the complete cases and missing cases on all variables included in the analyses; therefore, missing at random was assumed. To take into account the missing data and to avoid the reduction of sample size, full information maximum likelihood estimation in Mplus6.1 (Muthén & Muthén, 1998-2010) was used for regression analyses. Full information maximum likelihood uses all available data simultaneously to calculate the parameter estimates. This approach has been shown to produce less biased parameter estimates than listwise deletion or mean imputation (Enders, 2001) when data meet the assumption of missing at random.

Even so, we also ran our models without imputation ($n = 88$) using listwise deletion and only on the schools in which principals did not turnover ($n = 62$). Our assumption of missing at random was further supported by the finding that our substantive conclusions about the impact of the Balanced Leadership® program were not altered based on our sample approach (i.e., imputation or not, schools with turnover included or not). To illustrate this point, we report mean gain scores for the group of schools in which principals did not turn over ($n = 62$) before turning to regression findings that control for pretest status in the full group of 95 schools to estimate the causal impact of the program on the outcomes we measured. Our purpose is to illustrate that regardless of our approach to defining the sample, our substantive conclusions about the causal impact of the Balanced Leadership® program were unaltered. In addition, we believe some readers may find the presentation of gain scores in only the schools experiencing no principal turnover to be a relevant question that addresses the full impact of the program without personnel change. Because

turnover is an ecological reality, those interested in the impact of long-term professional development for leaders should know what net difference such training makes even when some principals change positions. Therefore, our final impact results are based on the full sample (n = 95). We turn next to the results of the program impact analyses described here.

Program Impact Findings

This section describes the findings of the analyses we designed to estimate the impact of the Balanced Leadership® program on principals. As we shall see in the results below, data presented in the tables and figures demonstrate that the Balanced Leadership® program delivered substantive and statistically significant positive treatment effects after the training was complete. In our evaluation framework, these results constitute evidence of learning, positive change in beliefs, and increases in some self-reported behaviors.

We begin with a presentation of descriptive statistics and a comparison between baseline attritors and active participants on these variables. As shown in Table 1, the average rate of poverty was just above 50% in our sampled schools. In addition, the data show no systematic attrition on demographic or achievement variables, with the participating schools (n = 95) being statistically equivalent to those who declined participation (n = 31) after learning of their random assignment status (treatment or control). Table 2 reports correlations among these variables. Not surprisingly, poverty was statistically related to baseline achievement scores in predictable directions. The results also show that poverty was more concentrated in smaller sample schools.

Next, in Table 3 we report treatment group means with control group means scale-by-scale for only the schools in which the principal did not turnover (n = 62). As described above, the unadjusted gain scores in Table 3 reflect the results reported by principals who remained in the same school from beginning to end of the study. The prescore means were taken from the first administration of the SLIS survey, before the treatment group principals attended training. The postscores were collected from principals on site in their schools approximately 2 months after BLPD training was completed. Because the assignment of treatment and control groups was random by design, group differences in subscales were not expected in the survey before intervention. Almost all scales failed to reject the null hypotheses of no group difference between treatment and control principals at baseline. Given the presence of only a few small but significant differences, we followed What Works Clearinghouse (WWC) standards (WWC, 2008) and controlled for pretest scores in all subsequent regression analyses.

Table 3. Gain Score *t* Test for Principal Survey First to Third Year—No Turnover (*N* = 62).

McREL Responsibility/Concept	Treatment (*n* = 32)		Control (*n* = 30)		*t* Value	*p* Value	Mean Difference
	Mean	SD	Mean	SD			
Parent and Community Involvement	0.438	0.831	−0.322	0.669	3.95	.000	.760**
Monitor/Evaluate	0.742	0.973	0.000	0.895	3.12	.003	.742**
School Norms for Differentiated Instruction	0.477	0.793	−0.186	1.005	2.89	.005	.663**
Principal Efficacy	0.581	0.769	−0.051	0.750	3.27	.002	.632**
Intellectual Stimulation	0.557	0.759	−0.011	0.821	2.83	.006	.568**
Principal Network Reflection	0.404	0.760	−0.159	0.870	2.72	.009	.563**
Teacher Collective Responsibility	0.469	0.718	−0.090	1.009	2.52	.014	.559*
Knowledge of Curriculum, Instruction, & Assessment	0.518	0.671	−0.025	0.603	3.35	.001	.543**
Discipline	0.396	0.894	−0.133	0.629	2.68	.009	.529**
Guaranteed and Viable Curriculum	0.328	0.774	−0.175	1.011	2.21	.031	.503*
Collaboration	0.418	0.653	−0.073	0.613	3.05	.003	.492**
Change Agent	0.398	0.672	−0.092	0.674	2.87	.006	.490**
Social Persuasion	0.508	0.776	0.025	0.531	2.87	.006	.483**
Ideals and Beliefs	0.406	0.603	−0.033	0.718	2.62	.011	.440*
Flexibility	0.313	0.713	−0.100	0.662	2.36	.022	.413*
Principal Leadership	0.369	0.552	−0.003	0.474	2.84	.006	.372**
Outreach	0.292	0.643	−.044	0.630	2.08	.042	.336*
Focus on Student Learning	0.458	0.974	0.175	0.783	1.27	.210	.283
Principal Trust in Teachers	0.113	0.498	−0.168	0.697	1.84	.071	.281†
Involvement in Curriculum, Instruction, & Assessment	0.406	0.923	0.128	0.753	1.30	.200	.278
Situational Awareness	0.284	0.533	0.018	0.580	1.88	.065	.266†
Visibility	0.302	0.814	0.078	0.917	1.02	.312	.224
Leading Change[a]	0.063	0.746	−0.161	0.992	1.01	.318	.224
Collective Efficacy	0.203	0.371	−0.019	0.578	1.81	.076	.222†
Principal's Support for Change	0.208	0.585	−0.011	0.719	1.32	.191	.219
Culture	0.208	0.784	0.022	0.625	1.03	.307	.186
Purposeful Community	0.331	0.831	0.192	0.706	0.71	.482	.139

[a]Gain score for Leading Change was calculated as the difference between second and third year given that it was not measured at the first year.
**Statistically significant effect at *p* < .01. *Statistically significant effect at *p* < .05. †Statistically significant effect at *p* < .10.

Turning to the gain scores in Table 3, treatment group means were consistently higher than control group means for all scales in the postsurvey. The mean gain scores of control schools ranged from −0.32 to 0.19 with an

overall mean of −0.05, whereas the range of treatment group mean gain scores was minimum 0.11 and maximum 0.74 with an overall mean of 0.38. More striking, the mean gain scores for the control group principals were negative for most scales, but the mean gains of the treatment group principals were positive across *all* indicators. Although it is only speculative, it may be the declines we observed in control schools were reflective of recession-era state and local budgets that were so seriously strained that some schools were threatened with the prospect of closure or emergency fiscal management.

More specifically, the gain score *t* test results showed a statistically significant mean difference between the reports of treatment principals and control school principals for *parent and community involvement*; *monitor/ evaluate*; *school norms for differentiated instruction*; *principal efficacy*; *intellectual stimulation*; *principal network reflection*; *teacher collective responsibility*; *knowledge of curriculum, instruction, and assessment*; *discipline*; *guaranteed and viable curriculum*; *collaboration*; *change agent*; *social persuasion*; *ideals and beliefs*; *flexibility*; *principal leadership*; and *outreach*. Marginal significance was shown for *principal trust in teachers, situational awareness*, and *collective efficacy.* This represents a statistically significant or marginally significant positive mean change on 20 of 27 (74%) BLPD indicators.

To reach stronger conclusions about the impact of the Balanced Leadership® program on principals, we conducted regression analyses (Table 4), controlling for presurvey scores and baseline school demographic variables (total enrollment, percent minority, percent free/reduced price lunch, and 2008 third-grade reading and math scaled score school means). As described earlier, we report both the unstandardized coefficients and Hedges' *g* effect sizes that are appropriate for small samples such as ours. In addition, we report both ITT and TOT effect sizes. The difference is that ITT estimates the net impact of the Balanced Leadership® program in the face of approximately 72% training attendance across treatment group principals (see Schroeder et al., 2012). In contrast, the TOT column estimates the impact one might expect if all principals had fully attended BLPD training. TOT estimates are useful because they offer a correction for the reduction in attendance that occurred because of treatment group principal noncompliance and turnover in the middle of Balanced Leadership training. In the case of 100% attendance, the ITT and TOT estimates would be equivalent.

In these analyses, treatment school principals showed positive and statistically significant differences from control school principals, on average, in terms of *principal efficacy, change agent, teacher collaboration, school norms for differentiated instruction, outreach, parent and community involvement, social persuasion, leading change, situational awareness,*

Table 4. Regression Analyses Results for Principal Survey at the Third Year[a]—Including Turnover (*N* = 95).

| Outcome Variable | Unstandardized Coefficients | | Effect Size (Hedges' *g* ITT) | Effect Size (Hedges' *g* TOT) | Significant Controls[b,c] |
	b	*SE*			
Principal Efficacy	0.424**	0.139	0.62	0.84	
Change Agent	0.337**	0.119	0.59	0.80	% Free/reduced price lunch[†]
Collaboration	0.348**	0.125	0.58	0.78	MEAP Reading[†]
School Norms for Differentiated Instruction	0.469**	0.169	0.57	0.77	
Outreach	0.384**	0.128	0.54	0.73	% Free/reduced price lunch[†]
Parent and Community Involvement	0.488**	0.159	0.54	0.73	% Free/reduced price lunch[†]
Social Persuasion	0.329**	0.126	0.53	0.72	
Leading Change[d]	0.476*	0.196	0.53	0.72	
Situational Awareness	0.327**	0.112	0.51	0.69	
Teacher Collective Responsibility	0.432*	0.175	0.49	0.66	
Intellectual Stimulation	0.400*	0.159	0.47	0.64	
Guaranteed and Viable Curriculum	0.363*	0.180	0.41	0.55	
Ideals and Beliefs	0.300*	0.136	0.41	0.55	% Free/reduce price lunch[†]
Monitor/Evaluate	0.345*	0.174	0.41	0.55	MEAP Reading[†]
Principal Trust in Teachers	0.237*	0.118	0.40	0.54	
Collective Efficacy	0.262*	0.104	0.36	0.49	
Principal Network Reflection	0.335*	0.164	0.35	0.47	% Minority[†]
Focus on Student Learning	0.231	0.166	0.32	0.43	
Knowledge of Curriculum, Instruction, & Assessment	0.224[†]	0.132	0.32	0.43	
Principal Leadership	0.150	0.099	0.27	0.36	
Flexibility	0.074	0.144	0.14	0.19	% Free/reduced price lunch**; MEAP Reading[†]
Discipline	0.090	0.154	0.13	0.18	

(continued)

Table 4. (continued)

Outcome Variable	Unstandardized Coefficients		Effect Size (Hedges' g ITT)	Effect Size (Hedges' g TOT)	Significant Controls[b,c]
	b	SE			
Culture	0.056	0.125	0.10	0.14	MEAP Reading[†]
Principal's Support for Change	0.042	0.146	0.09	0.12	
Visibility	0.081	0.175	0.09	0.12	
Involvement in Curriculum, Instruction, & Assessment	0.043	0.149	0.05	0.07	% Free/reduced price lunch*; Total School Enrollment[†]
Purposeful Community	0.039	0.144	0.05	0.07	%Free/reduced lunch[†]; MEAP Reading*

Note. MEAP = Michigan Educational Assessment Program; ITT = intent to treat; TOT = treatment on the treated.
[a]The number of schools included in the analyses was 95. [b]The number of total enrollment in school, % of minority, % of free and reduced lunch, 2008 MEAP Reading and Math scaled scores for third grade, and the corresponding baseline (Year 1) mean score for each measure were included in the regression models as covariates. [c]All baseline (Year 1) mean scores were statistically significant at *p* < .05 except for Guaranteed and Viable Curriculum. [d]Given that Leading Change was not measured at Year 1, Instructional Leadership was used as the corresponding baseline covariate.
**Statistically Significant effect at *p* < .01. *Statistically significant effect at *p* < .05. [†]Statistically significant effect at *p* < .10.

teacher collective responsibility, intellectual stimulation, guaranteed and viable curriculum, ideals and beliefs, monitor/evaluate, principal trust in teachers, collective efficacy, and *principal network reflection* at the .05 significance level or lower. The mean difference between the treatment and control groups in *knowledge of curriculum, instruction*, and *assessment* was marginally significant with a *p* value slightly more than .05.

Next, we compared the treatment effects in the gain score analysis for schools with no principal turnover (*n* = 62) to the regression-adjusted results obtained in the full sample of 95 participating schools. Notably, although the ascending orders of magnitude are not identical, patterns of statistical significance are the same for 19 outcomes while four that were significant in the gain score analysis were only marginally or nonsignificant in the regression analyses. In sum, these comparisons suggest that the impact of the program on treatment school principals for the outcomes measured is generally positive.

Although the strong positive treatment effects do not appear to be a function of chance, we were still concerned with the possibility of inflated Type I error rates, which can occur when multiple comparisons are made. Therefore,
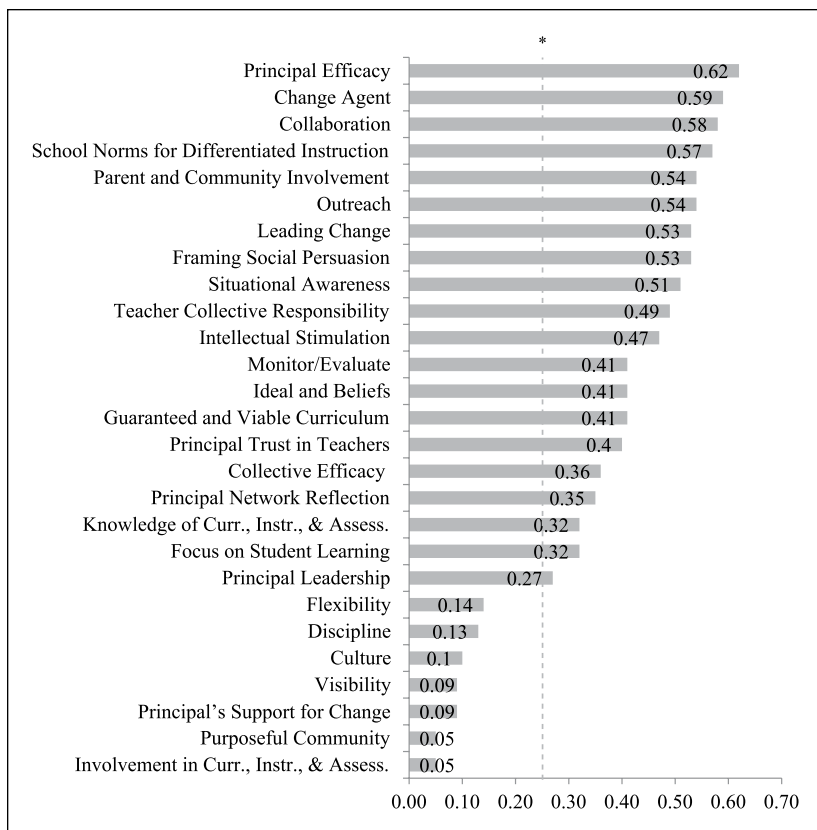
Figure 1. Hedges' *g* effect size for BLPD treatment effects on principals at third year (ITT) (*n* = 95)*.
Note. *Effect sizes of .25 or greater should be considered a substantively significant effect (Rosenthal, 2004; Rosenthal, Rosnow, & Rubin, 2000; What Works Clearinghouse, 2008).

we decided to interpret the substantive importance of our findings according to WWC Standards (WWC, 2008) for experimental research. Specifically, according to WWC, in experimental research treatment effects of .25 or greater should be considered substantively meaningful regardless of *p* values. To illustrate the effect size differences between treatment and control principals, we created a figure to complement the findings of the regression analyses in Table 4. Figure 1 visually displays the more conservative ITT program impact estimates relative to a vertical line marking the .25 substantive significance threshold identified by WWC. Figure 1 demonstrates a clear pattern of

positive treatment effects for the Balanced Leadership® program with more than three quarters of the proximal indicators exceeding the WWC standard of .25 for indication of substantively important treatment effects (note that the pattern of substantively significant effects is the same for the TOT estimates reported in Table 4). A total of 21 of the 27 measures exceed the WWC effect size standard with ITT effect sizes ranging from .32 to .64. Notably, for the effects that do not exceed .25, all differences favor the treatment positively. We therefore conclude that these positive effects are attributable to the Balanced Leadership® program, the only variable manipulated in this experimental study.

To advance knowledge about the types of impacts professional development for school principals may have, we also wanted to know whether there were patterns in the magnitude of outcomes across the domains of principal efficacy beliefs, norms for teacher practice (collaboration and differentiated instruction), school climate, and principal leadership. The first major finding is that the largest impact of the program as measured by the Hedges' *g* effect size was on principal efficacy for instructional leadership (ITT ES =.62). This means that while principals in the treatment and control groups entered the study at baseline with no significant differences in their sense of efficacy, after receiving the Balanced Leadership® program, those in the treatment group scored .62 standard deviations higher in their sense of capability for bringing about improvements in instruction and learning in their schools. Treatment principals also uniformly reported higher than control principals after the training in the area of school norms for effective teacher practice. More specifically, treatment school principals perceived higher levels of school wide collaboration among teachers (ITT ES =. 58) and stronger norms for differentiated instruction (ITT ES = .57). In contrast, the ITT effect sizes in the domain of principal leadership, while all positive, ranged from high and substantively significant to low and not substantively significant. For example, the principal leadership skill of being a change agent (e.g., *I create the conditions necessary for change to happen*) had an ITT effect size of .59. In contrast, the instructional leadership behavior of involvement in curriculum, instruction, and assessment had an ITT effect size of only .05.

To examine within domain differences for school climate and principal leadership more carefully, we prepared two tables that rank order the effect sizes from high to low. Table 5 ranks all effect sizes that were substantively significant based on WWC standards (i.e., ES ≥ .25). Next, Table 6 ranks the effect sizes that were less than .25 standard deviations (i.e., nonsubstantively significant). These tables present the name of the outcome construct, the associated effect size, and a sample item representing the types of questions principals answered.

Table 5. Hedges' *g* Effect Size: ≥.25.

Subscale	Effect Size	Sample Item
Principal Efficacy	.62	My leadership skills are currently sufficient to improve instruction in this school in ways that foster high levels of student learning.
		I have what it takes now to lead to high levels of student learning in my school.
Change Agent	.59	I create the conditions necessary for change to happen.
Collaboration	.58	Teachers in this school work collectively to plan school improvement.
School Norms for Differentiated Instruction	.57	Teachers in this school use a wide range of assignments, materials, or activities matched to students' needs and skill levels.
Parent and Community Involvement	.54	We encourage feedback from parents and the community.
Outreach	.54	I am a strong advocate for our school to the community at large.
Leading Change	.53	I have implemented one or more major change initiatives in our school in the last 12 months.
Framing & Social Persuasion	.53	I regularly frame the work around school improvement as doable.
Situational Awareness	.51	I am aware of both what is and what is not running smoothly in our school.
Teacher Collective Responsibility	.49	Teachers in this school are responsible to help each other do their best.
Intellectual Stimulation	.47	We regularly have discussions about current research and theory.
Monitor/Evaluate	.41	I continually monitor the effectiveness of our curriculum.
Ideal and Beliefs	.41	My behavior is consistent with my beliefs and ideals regarding schools, teachers, and learning.
Guaranteed and Viable Curriculum	.41	There is constant focus on the goals for learning the essential curriculum.
Principal Trust in Teachers	.40	I can rely on the teachers in this school.
Collective Efficacy	.36	Teachers in this school believe that every child can learn.
Principal Network/ Reflection	.35	I regularly reflect on the focus of change for our school and my management of it.
Knowledge of Curriculum, Instruction, & Assessment	.32	I am very knowledgeable about effective instructional practices
Focus on Student Learning	.32	When making important decisions, the school always focuses on what's best for student learning.
Principal Leadership	.27	I set high standards for teaching.

Table 6. Hedges' g Effect Size: <.25.

Subscale	Effect Size	Sample Item
Flexibility	.14	I encourage people to express opinions that may be contrary to my own.
Discipline	.13	I protect instructional time from interruptions.
Culture	.10	I promote cohesion among teachers and staff.
Visibility	.09	I make systematic and frequent visits to classrooms.
Principal's Support for Change	.09	I encourage teachers to try new methods of instruction.
Involvement in Curriculum, Instruction, & Assessment	.05	I am directly involved in helping teachers design curricular activities in their classrooms.
Purposeful Community	.05	I focus on what is most important for the success of students in this school.

A careful reading of these tables revealed several patterns. First, aside from the principal efficacy measure, the constructs on which the professional development had substantively significant impacts generally represent broad, school-level phenomena. These include, for example,

- Teachers in this school work collectively to plan school improvement.
- We encourage feedback from parents and the community.
- There is constant focus on the goals for learning the essential curriculum.
- I set high standards for teaching.

In contrast, the constructs in Table 6 where the nonsubstantively significant effects are reported have a different orientation. More specifically, the areas where the professional development appears to have had smaller and nonsubstantively significant effects regard principals' reports on whether they perform specific behaviors, many of which would involve them more directly in teachers' work. Sample items in this area include the following:

- I am directly involved in helping teachers design curricular activities in their classrooms.
- I encourage teachers to try new methods of instruction.
- I make systematic and frequent visits to classrooms.

Thus, the BLPD program appears to have significantly affected principals' sense of efficacy and most measures of school climate including norms for differentiated instruction among teachers and some leadership behaviors, particularly the management of change. In contrast, treatment principals were less likely on average to report substantively significant increases in specific behaviors, several of which involve their close work with teachers. We turn next to a discussion of these findings.

Discussion

This 5-year evaluation provides evidence of substantive growth in principals' self-reported learning and behavior as a result of participating in McREL's Balanced Leadership® program. Overall, we are encouraged that a research-based leadership training program has produced effects of the magnitude we discovered through this investigation. As our findings indicated, however, the most important lessons in this study seem to be about differences in the ways principals did and did not grow as a function of the training.

One of the questions we examined regarded whether principals experienced greater growth on average in some domains than in others. As shown in Table 5 and Figure 1, the largest areas of impact were principal efficacy for instructional leadership (ES = .62), school norms for teacher collaboration and differentiated instruction (ES = .58 and .57, respectively), and principals' reports of their ability to manage change (ES = .59). Looking at the largest program impact areas, one conclusion is that participants left the program with greater beliefs in their ability to lead instructional improvement, a heightened sense of awareness regarding teachers' work together and use of particular teaching practices, and enhanced knowledge about how to lead changes necessary for school improvement. Such interpretations would be consistent with major outcomes stated by the program developer.

Another pattern that emerged was that principals' perceptions of learning from the program tended to increase more than their reports of specific leadership behaviors and actions that involved them directly with teachers. A good comparison that illustrates this pattern involves principals' reported *knowledge* of curriculum, instruction, and assessment (ITT ES = .32; Table 5) versus principals' reported *involvement* in curriculum instruction and assessment (ITT ES = .05; Table 6). By their own reports, principals grew over six times more in their knowledge than involvement in these areas. Similarly, principals also reported higher on other measures containing items that did not involve them directly in teachers' work (e.g., "I have what it takes to lead to high levels of

student learning in my school" and "Teachers in this school work collectively to plan school improvement") than those probing whether leaders were directly involved in the detailed professional work of teaching (e.g., "I am directly involved in helping teachers design curricular activities in their classrooms" and "I encourage teachers to try new methods of instruction").

There are several possible explanations for these findings. First, it may be that the Balanced Leadership® program placed more emphasis on content and experiences that build efficacy for leading improvement, and understanding the changes and group processes this requires, than on how principals can or should be involved directly with teachers in their daily instructional work. Indeed, given that principals usually lead schools with teachers who collectively have advanced content knowledge in a wide variety of subjects, this could make sense. It may also be that the elementary school leaders we studied conceived the importance and impact of their role as providing a positive climate for change and necessary instructional supports while allowing teacher autonomy regarding the enactment of teaching. Another possibility is that for this study Balanced Leadership® training was offered only to principals and that, by design, teachers did not participate in the training. Given the literature on distributed leadership (e.g., Spillane, 2012) and the role of teacher leadership teams (Marks & Printy, 2003), others may wish to investigate the outcomes of research-based professional development that brings school leaders and teacher leadership teams together for purposes of instructional improvement.

A limitation of this study is that it examines only principals' perspectives. Teachers' perspectives and student outcomes are also important and, because we did not include those perspectives in this article, we did not triangulate them with principals' perceptions. Readers should understand, therefore, that the results of this article are, in fact, limited to principals' perceptions.

Even so, the findings reported here raise other intriguing possibilities. For example, future researchers may wish to investigate whether nonsignificant growth in principals' direct involvement with teachers is capable of promoting student achievement when professional development leads to growth in broader, school-level areas such as those identified here. Although we cannot answer this question definitively, we suspect the answer is "no" in the short-term as we have reported elsewhere that the BLPD training did not affect student achievement (Jacob et al., 2015) during the timeframe of the study. Even so, the program did have a large and significant impact on principals' sense of efficacy for instructional leadership.

Conclusion

Findings from this randomized controlled trial show that the Balanced Leadership® program yielded several substantively important and positive effects on principals' perceptions of efficacy, leadership, climate, and norms for teacher practice. This conclusion is further supported by companion work showing that BLPD training is implemented with a high degree of fidelity across and within treatment sites (Schroeder et al., 2012). Together, these study results across 27 outcomes suggest that, on average, the treatment principals perceptions were positively influenced by participation in BLPD.

The most intriguing findings from this study appear to involve the types of outcomes the program did and did not produce. While we are strongly encouraged that principals' sense of efficacy for instructional leadership and beliefs in their ability to manage complex change increased because of the BLPD program, we are nevertheless concerned about some differences that emerged. Specifically, the professional development studied appeared to cause generally strong growth in principals' perceptions in broad areas of school climate. In contrast, principals who received the training were not as likely to report substantively significant growth in the practices that involved them directly in teachers' work around curriculum, instruction, and assessment. Thus, while we have some encouraging findings, we believe there is exciting work remaining to determine what types of professional development are needed to develop leaders who can improve teaching and learning. The findings of this study, in conjunction with emerging research on principal professional development programs that involve leaders working more directly with teachers (e.g., Darling-Hammond et al., 2009; Grissom & Harrington, 2010; Hamilton et al., 2012), may have the potential to transform future principal professional development practices in ways that may lead to significant improvements for teachers and students. In fact, future research should focus more on the actual behavioral changes that principals make as a result of professional development opportunities and solicit teachers' input regarding these changes. In summary, the implications of our findings include the need for principals to put into action what they learn during formal professional development and the possible need to involve teachers in leadership professional development opportunities so that they may also influence positive change in their schools.

Appendix A
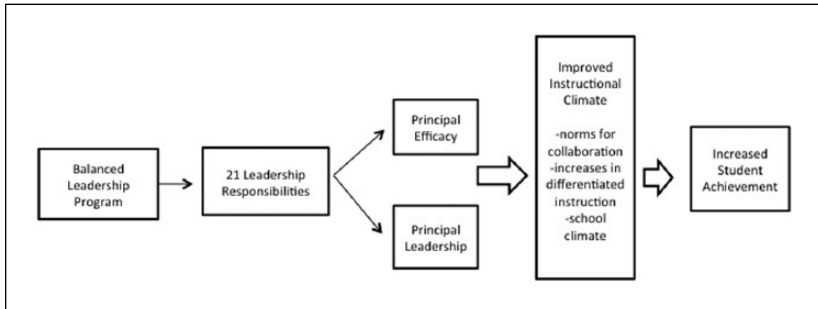
Hypothesized pathways of influence connecting the balanced leadership program to student achievement.

Appendix B

Balanced Leadership® Program School Principal Leadership Responsibilities, Alphabetical Listing

1. Affirmation—Recognition of accomplishments and acknowledgment of failures
2. Change agent—Tendency toward bringing about change
3. Communication—Established with and between staff and students
4. Contingent rewards—Recognizes individual successes and accomplishments
5. Culture—Fosters shared beliefs and community
6. Discipline—Protecting teachers' time from distractions
7. Flexibility—Adapts to current conditions and is comfortable with strife
8. Focus—Emphasizes clear goals
9. Ideals and beliefs—Communicates and demonstrates strong education beliefs
10. Input—Involves teachers in decision making and policy formation
11. Intellectual stimulation—Keeps faculty aware of current practices/theories in education
12. Involvement in curriculum, instruction, and assessment—Directly involved in design and implementation
13. Knowledge of curriculum, instruction, and assessment—Aware of best practices in these areas

14. Monitoring/evaluating—Monitors effectiveness of teaching and student learning
15. Optimizer—Inspires others and supports innovation
16. Order—Establishes structure, routines and procedures
17. Outreach—Advocates for the school to all stakeholders
18. Relationships—Shows personal regard and awareness toward teachers and staff
19. Resources—Provides materials and professional development for staff
20. Situational awareness—Aware of and sensitive to school environment and events
21. Visibility—Interacts with constituents

Authors' Note

Declaration of Conflicting Interests

Funding

Notes

1. The Interstate School Leadership Licensure Consortium standards are now in draft form with the previous version released in 2008.
2. Although the overarching goal of this large-scale, randomized control trial included assessing the training fidelity, and causal impacts of the McREL program on teacher perceptions, student achievement, and educator turnover, the present study is focused explicitly on how the program affected leaders' reports of change. See Jacob et al. (2015) for a report on other outcomes.

3. One coauthor of the article has published research that is cited in the BLPD program materials and spent a day with members of McREL over a decade ago speaking about this research. The BLPD program, however, was developed and deployed by McREL without any other input from this co-author. Furthermore, this coauthor has no ownership of the program and does not gain financially from the success or failure of the program or its evaluation results. A conflict of interest management plan was in place to avoid any potential compromise to the integrity of the study and its findings. The primary role of all coauthors was only to evaluate the program.
4. Note that because public data on the demographic characteristics of individual schools were available to inform our selection and sample stratification processes in ways that were designed to avoid systematic differences in the professional contexts in which the treatment and control school principals worked, we discuss our recruitment of principals in terms of the schools they served.
5. Although the experimental groups did not differ on the stratifying variables, we also tested for differences in teacher experience and certification between the treatment and control groups. Our findings revealed that there were no preexisting differences on the teacher experience variable. As such, there was no need to control for this. For teacher certification, we found a substantively small difference between the rates of teacher certification between the treatment and control groups. However, the teacher certification variable was also statistically unrelated to any of the 27 outcomes studied here so we did not include it.
6. Between baseline and the third wave of data collection we lost three additional treatment schools and four additional control schools, for a final sample of 88 schools (40 treatment and 48 control), but because we had some data available, we impute outcome data for these schools.

References

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, *50*, 341-358.

Argyris, C., & Schön, D. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass.

Argyris, C., & Schön, D. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison Wesley.

Berebitsky, D., Goddard, R., & Carlisle, J. (2014). An examination of teachers' perceptions of principal support for change and teachers' collaboration and communication around literacy instruction in reading first schools. *Teachers College Record*, *116*(4), 1-28.

Bloom, H. S. 1984. Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, *8*(2), 225-246.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (No. w17803). Cambridge, MA: National Bureau of Economic Research.

Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Council of Chief State School Officers. (2014). *2014 ISLLC standards*. Washington, DC: Author.

Darling-Hammond, L., Meyerson, D., LaPointe, M., & Orr, M. T. (2009). *Preparing principals for a changing world: Lessons from effective school leadership programs*. San Francisco, CA: Wiley.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Journal of Educational and Psychological Measurement*, *61*, 713-740.

Fullan, M. (2011). *Change leader: Learning to do what matters most*. San Francisco, CA: Jossey-Bass.

Goddard, R. D. (2001). Collective efficacy: A neglected construct in the study of schools and student achievement. *Journal of Educational Psychology*, *93*, 467-476.

Goddard, R. D., Goddard, Y. L., Kim, E. S., & Miller, R. J. (2015). A theoretical and empirical analysis of the roles of instructional leadership, teacher collaboration, and collective efficacy beliefs in support of student learning. *American Journal of Education*, *121*, 501-530.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher*, *33*, 3-13.

Goddard, R. D., & LoGerfo, L. (2007). Measuring emergent organizational properties: A comparison of the predictive validity and intergroup variability of self vs. group referent perceptions. *Educational and Psychological Measurement*, *67*, 845-858.

Goddard, R. D., Salloum, S., & Berebitsky, D. (2009). Trust as a mediator of the relationships between academic achievement, poverty and minority status: Evidence from Michigan's public elementary schools. *Educational Administration Quarterly*, *45*, 292-311.

Goddard, Y. L., Goddard, R. D., & Kim, M. (2015). School instructional climate and student achievement: An examination of group norms for differentiated instruction. *American Journal of Education*, *122*, 111-131.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*, *109*, 877-896.

Grissom, J. A., & Harrington, J. R. (2010). Investing in administrator efficiency: An examination of professional development as a tool for enhancing principal effectiveness. *American Journal of Education*, *116*, 583-612.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.

Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, *33*, 329-351.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, *4*, 221-239.

Hamilton, L. S., Engberg, J., Steiner, E. D., Nelson, C. A., & Yuan, K. (2012). *Improving school leadership through support, evaluation, and incentives: The Pittsburgh Principal Incentive Program*. Santa Monica, CA: RAND.

Hausman, C. S., & Goldring, E. B. (2001). Sustaining teacher commitment: The role of professional communities. *Peabody Journal of Education*, *76*(2), 30-51.

Jackson, B. L., & Kelley, C. (2002). Exceptional and innovative programs in educational leadership. *Educational Administration Quarterly*, *38*, 192-212.

Jacob, R., Goddard, R. D., Kim, M., Miller, R. J., & Goddard, Y. L. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, *37*, 314-332.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of American Society of Training Directors*, *13*(3), 21-26.

Kirkpatrick, D. L. (1994). *Evaluating training programs*. San Francisco, CA: Berrett-Koehler.

Knapp, M. S., Copland, M. A., & Talbert, J. E. (2003, February). *Leading for learning: Reflective tools for school and district leaders* (Research Report). Seattle, WA: Center for the Study of Teaching and Policy.

Lee, V. E., & Smith, J. B. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. *American Journal of Education*, *104*, 103-147.

Leithwood, K., Riedlinger, B., Bauer, S., & Jantzi, D. (2003). Leadership program effects on student learning: The case of the greater New Orleans school leadership center. *Journal of School Leadership*, *13*, 707-738.

Levine, A. (2005). *Educating school leaders* (Education Schools Project). New York, NY: Teacher's College, Columbia University. Retrieved from http://www.edschools.org/pdf/Final313.pdf

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2009). *How does leadership affect student achievement? Results from a national survey*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*, 370-397.

Miller, R. J., Goddard, R. D., Kim, M., Jacob, R., Goddard, Y. L., & Schroeder, P. (2013). Balanced Leadership Professional Development training effects on principals' school leadership practice: Teacher perceptions from a randomized control trial. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus user's guide: Statistical analysis with latent variables* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Owings, W. A., Kaplan, L. S., & Nunnery, J. (2005). Principal quality, ISLLC standards, and student achievement: A Virginia study. *Journal of School Leadership*, *15*, 99-119.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*, 635-674.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York, NY: Russell Sage Foundation.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. New York, NY: Cambridge University Press.

Schroeder, P., Kim, E. S., Goddard, R. D., Miller, R. J., Madsen, J., & Goddard, Y. L. (2012). *Assessing the fidelity of a principal professional development program: Design, delivery, and receipt*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Sheeler, M. C., Bruno, K., Grubb, E., & Seavey, T. L. (2009). Generalizing teaching techniques from university to K-12 classrooms: Teaching preservice teachers to use what they learn. *Journal of Behavioral Education*, *18*, 189-210.

Spillane, J. P. (2012). *Distributed leadership* (Vol. 4). Hoboken, NJ: Wiley.

U.S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. Washington, DC: Author.

Waters, J. T., Marzano, R. J., & McNulty, B. A. (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.

What Works Clearinghouse. (2008). *What Works Clearinghouse: Procedures and standards handbook* (Version 2.0). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Author Biographies

Robert James Miller serves as a senior evaluator in the Department of Federal and State Programs for the Edgewood Independent School District in San Antonio, TX. He has extensive experience conducting program evaluation and analysis of large-scale education databases. His work focuses on education policy, organization theory, and the analysis of school effectiveness.

Roger D. Goddard is a professor and the Novice G. Fawcett Chair in Educational Administration at The Ohio State University. His research interests include organizational theory, social psychology, school leadership, experimental design, and hierarchical linear modeling.

Minjung Kim received her PhD from Texas A&M University (2012) and is currently a postdoctoral fellow of the Psychology Department at the University of South

Carolina. Her research interests include longitudinal data analysis using multilevel models and structural equation models applied to social science research. She is also interested in investigating the differential effect using the regression mixture modeling approach.

Robin Jacob is a research assistant professor with a joint appointment at the School of Education and the Institute for Social Research at the University of Michigan. Much of her research centers on the evaluation of educational interventions, with a special focus on how policies and programs can affect instructional quality and outcomes in elementary schools. She has extensive experience conducting evaluations of education reform initiatives, measuring educational outcomes, and analyzing student achievement and other outcome data.

Yvonne Goddard is a visiting associate professor at The Ohio State University, and she was a principal investigator for the School Leadership Improvement Study. Her research focuses on teacher collaboration and effective instructional practices for academic improvement. Her most recent research focuses on teachers' use of differentiated instruction, which she has successfully linked to collective efficacy, student achievement, and principals' instructional leadership.

Patricia Schroeder received her PhD from Texas A&M University (2013). She brings a strong background in qualitative analysis applied to program evaluation and a wealth of "real-world" experience in educational administration and school leadership as a former school administrator. Her research interests focus on the design, delivery, and receipt of professional development programs received by school leaders.