

How Does District Principal Evaluation Affect Learning-Centered Principal Leadership? Evidence From Michigan School Districts

MIN SUN and PETER YOUNGS

Michigan State University, College of Education, East Lansing, Michigan, USA

This study used Hierarchical Multivariate Linear models to investigate relationships between principals' behaviors and district principal evaluation purpose, focus, and assessed leadership activities in 13 school districts in Michigan. The study found that principals were more likely to engage in learning-centered leadership behaviors when the purposes of evaluation included principal professional development, school restructuring, and accountability; when the focus of evaluation was related to instructional leadership; and when evaluation addressed leadership in school goal setting, curriculum design, teacher professional development and evaluation, and monitoring student learning. The findings from this study have implications for improving district evaluation policies and practices.

INTRODUCTION

In the United States, the recent focus on accountability policies in K–12 education has concentrated attention on the role of school leaders in improving student performance (Stein & Spillane, 2005; Mangin, 2007). The federal No Child Left Behind (NCLB) legislation and state accountability systems have placed greater responsibility on principals for initiating reforms and enhancing student learning (Leithwood, Louis, Anderson, & Wahlstrom, 2004; Marks & Nance, 2007). Further, educators, researchers, and policymakers are increasingly interested in whether and how principal

The authors would like to thank the Michigan State University Educational Policy Center for financial support and Wang Jun Kim for assistance with survey administration.

Address correspondence to Min Sun, Ph.D., Michigan State University, College of Education, 203 Erickson Hall, East Lansing, MI 48824, USA. E-mail: sunmin@msu.edu

evaluation systems can influence principals' leadership behaviors and school improvement efforts (Portin, Feldman, & Knapp, 2006).

In the United States, districts have traditionally used straightforward approaches to evaluate principals such as checklists of behaviors, an annual visit by the district superintendent, and/or a formal review related to contract renewal. In most districts, these activities have been intended to maintain some accountability for principals' work; to justify hiring, reassignment, or termination decisions; or to inform renewal of school leaders' contracts. At the same time, little evidence exists of the validity or reliability of principal evaluation systems that feature these approaches (AERA, APA, & NCME, 1999). In addition, these evaluation practices have focused primarily on principals' personal characteristics as opposed to the activities in which they engage or the outcomes of their work (Heck & Marcoulides, 1992; Portin, Feldman, & Knapp, 2006). Oftentimes, the evaluation practices have been established based on little understanding of how principal evaluation is related to leadership practices. Further, the extant research literature provides little guidance regarding whether or how the purpose or focus of principal evaluation is related to learning-centered leadership behaviors. As a result, educators, researchers, and policymakers all recognize a need for research on the mechanisms by which district evaluation affects leadership and how evaluation strategies can be used to promote effective leadership.

According to recent theoretical work (Portin, Feldman, & Knapp, 2006), under certain conditions principal evaluation is likely to be associated with efforts by school leaders to establish rigorous goals for student learning, enact demanding curricula, and closely monitor and support teachers' instructional practices. Moreover, a series of leadership activities that could improve student learning have been identified in prior research (Murphy, Goldring, Cravens, Eliot, & Porter, 2007). But little empirical research exists on how the purpose and focus of principal evaluation systems, or the leadership activities they assess, are related to these learning-centered leadership behaviors. The purpose of this article is to address this shortcoming in the literature by investigating relationships between principals' behaviors and principal evaluation purpose, focus, and assessed leadership activities in 13 school districts in Michigan. In addition, this article examines several ways in which district evaluation policy could increase the likelihood of promoting learning-centered leadership. In particular, we use Hierarchical Multivariate Linear models to measure the relationship between aspects of principal evaluation and the outcomes of interest, learning-centered leadership behaviors.

In the first section of this article, we briefly review research on school leadership and principal evaluation. The second section explicates the theoretical framework that informed our research design and data analysis. In the third section, we describe our sample, measures, and models. The fourth section presents the main findings from our analyses while the fifth

section discusses how these findings contribute to the research literature. Finally, we conclude by considering some implications for further research as well as some limitations of this study.

RESEARCH ON PRINCIPAL LEADERSHIP AND PRINCIPAL EVALUATION

This section first examines research on principal leadership and principal evaluation. Our review of research on school leadership is based on recent comprehensive reviews by Portin, Feldman, and Knapp (2006) and Murphy et al. (2007) and focuses on empirical studies that included student achievement as an outcome measure or intermediate factors that research has indicated are related to achievement. Our review of research on principal evaluation addresses empirical studies from the 1980s, as well as several studies from the 1990s and 2000s. This review is limited by the paucity of empirical research in this area. Finally, this section discusses the limited empirical research on how principal evaluation is related to principal leadership.

Research on Principal Leadership

Research on the principalship indicates that school leaders can impact student achievement indirectly through their influence on learning goals, demanding curricula, effective pedagogy, teacher commitment, accountability, and teacher professional community and relational trust (Hallinger & Heck, 1996; Leithwood et al., 2004; Nettles & Harrington, 2007). A number of studies indicate that rigorous learning goals are associated with higher levels of student achievement (Betts & Grogger, 2003; Hallinger, Bickman, & Davis, 1996; Lee & Smith, 1996; Louis & Marks, 1998; Marks & Printy, 2003). When principals work with teachers to establish and maintain such goals for all students, they are likely to help their schools improve student performance. Similarly, research provides evidence that instruction that focuses on enacting ambitious curricula is related to increased student learning (Gamoran, Porter, Smithson, & White, 1997; Knapp, Shields, & Turnbull, 1992; Wong, Hedges, Borman, & D'Agostino, 1996). Thus, school leaders are likely to promote increases in student achievement when they ensure that all students have access to demanding subject matter content.

Effective pedagogy, or quality instruction, is defined as teaching practices that maximize student engagement and learning. Principals can support such instruction by connecting teachers to professional development that focuses on instruction and student outcomes, includes opportunities for assistance, and is sustained and continuous (Cohen & Hill, 2001; Desimone, Porter, Garet, Suk Yoon, & Birman, 2002); and by promoting coherence in professional development and instruction (Newmann, Smith, Allensworth, &

Bryk, 2001). In addition, school leaders can enhance student learning by strengthening teacher commitment and job satisfaction (Marks & Printy, 2003; Ross & Gray, 2006; Snyder & Ebmeier, 1992).

Principals can also promote high levels of student achievement when they integrate internal and external accountability systems and hold all teachers accountable for student learning (Newmann, King, & Rigdon, 1997; Murphy et al., 2007). This involves communicating expectations to teachers and providing them with consistent feedback on classroom and school performance. Finally, research indicates that principals can have positive effects on student learning when they promote professional community among teachers and relational trust between themselves and teachers and between school staff and students and their families (Bryk & Schneider, 2002; Louis & Marks, 1998; Marks & Printy, 2003).

Research on Principal Evaluation

In the 1980s several researchers investigated state and district approaches to evaluating school leaders and holding them accountable for their work. In one study, for example, Harrison and Peterson (1986) surveyed superintendents and principals in order to examine the implementation of a statewide approach to principal evaluation. While the evaluation system was highly specified, implementation was inconsistent across districts due to variations in how leaders' performance was sampled, how the outcomes of their work were assessed, and the ways in which feedback was communicated (Harrison & Peterson, 1986). In a second study, Garrett and Flanigan analyzed a principal evaluation system in one South Carolina district that featured data on school leaders obtained from teachers, parents, and district administrators. The authors considered challenges to enacting such an approach to evaluation as well as its positive effects on principals' behaviors (Garrett & Flanigan, 1991).

Other research in the 1980s examined ways in which districts hold school leaders accountable and both support and constrain their work. In one study, Peterson (1984) investigated the ways in which district administrators in 59 districts shaped the work of 113 elementary school principals through both tight and loose controls. While the controls used across districts included a combination of hierarchical, social, and environmental elements, the patterning of control and the degree of principal autonomy in each school were strongly influenced by both the social status of the school and the size of the district (Peterson, 1984). In another study, Murphy and colleagues (1987) analyzed the administrative control processes and activities used in 12 instructionally effective school districts in California. Based on interviews with superintendents in the 12 districts, the authors reported high levels of district-level control of principal behaviors and school activities and noted the crucial role played by superintendents in linking schools to district administrators (Murphy, Hallinger, Peterson, & Lotto, 1987).

More recently, a number of studies have focused on the perceptions of and responses to district principal evaluation policies on the part of school and district leaders. In one study, Davis and Hensley (1999) interviewed 14 principals and six superintendents in northern California in order to examine the extent to which the groups held similar views about principal evaluation. Their findings indicated that the evaluation formats and processes often varied from one district to another and that principals and superintendents frequently held different perspectives about evaluation purposes and benefits. Most principals felt their evaluations were not useful and were affected by political forces beyond their control. A second study by Thomas, Holdaway, and Ward (2000) investigated principal evaluation policies and practices used by school systems in Alberta, Canada. Their data included superintendent and principal surveys, evaluation documents, and interviews with 10 superintendents and 10 principals. Similar to Davis and Hensley, their results revealed substantial diversity across jurisdictions in terms of principal evaluation practices, several constraints upon effective evaluation, and significant differences between the opinions of superintendents and principals regarding evaluation purposes.

More recently, in a survey of principals from throughout the United States, Reeves (2005) found that few school leaders viewed principal evaluation as affecting their motivation or performance. Most principals in this study felt that the criteria used to evaluate them were poorly specified, that the evaluation process had few consequences, and that it did not provide useful feedback to them. In another study, Goldring and colleagues (2008) examined the content and use of principal evaluation instruments in 65 urban school districts in more than 40 states. The researchers found that the instruments in their study primarily focused on school leaders' efforts to establish rigorous learning goals, promote teacher professional community, and hold school staff accountable for student learning. At the same time, the instruments sampled paid relatively less attention to principals' efforts to implement ambitious curricula or monitor instruction (Goldring et al., 2008). The authors also reported that most principal evaluation systems were used for formative purposes and few employed evaluation criteria based on standards or evidence.

In a fifth study, Kimball and Pautsch (2008) investigated the enactment of standards-based principal evaluation systems in two large, urban districts. Based on interviews with seven principal evaluators and 14 school leaders in each district, the study found that the two districts differed in how they defined and assessed instructional leadership. One district focused on an instructional strategy that stressed student engagement while the other emphasized the development of school-based professional learning communities. But neither district clearly communicated the role evaluation would play in overall school improvement strategies. Further, principal evaluators held multiple responsibilities within these districts in addition to

their evaluation roles and struggled to complete comprehensive evaluations. As a result, neither district implemented the process in a way that maximized rating consistency.

These studies provide evidence that principal evaluation policies and practices often vary across school districts and that school leaders and district administrators typically have divergent views of evaluation purposes. While the Goldring et al. (2008) and Kimball and Pautsch (2008) studies suggest that district evaluation policies have begun to focus in recent years on student content standards, accountability, and school change, none of the studies described above examined the relationship between principal evaluation and school leaders' behaviors. In fact, a thorough review of the research literature found only a few empirical studies that had considered this relationship.

In one study, Smith, Munter, and Katterfield (2008) assessed the extent to which "messages" regarding district expectations for math instruction were associated with principals' levels of math leadership content knowledge, including their views of high-quality math instruction and how they supported and evaluated math teachers' instruction. Data came from surveys and interviews of 120 teachers in 38 schools across four districts, surveys and interviews of the principals in these schools, interviews with district administrators responsible for principal evaluation, and documentation of the principal and teacher evaluation procedures in each district. Findings indicated that principals in districts whose reform goals were unaligned with leadership assessment systems were more likely to focus on surface aspects (e.g., student test-score data) in evaluating teachers than on promoting high-quality instruction. Further, principals who felt their district's evaluation process was consistent with the district's goals for improving instruction were more likely to strengthen their leadership content knowledge by participating in professional development or seeking advice from individuals with math expertise.

In a second study, Kimball, Milanowski, and McKinney (2007) conducted a randomized design study of principal evaluation and outcomes for school leaders in a large school district in the western U.S. For the study, 44 principals were randomly assigned to a new standards-based performance evaluation system and 44 principals were evaluated under the district's old system.¹ The researchers found that use of the new evaluation system seemed to be associated with several noteworthy outcomes, but few of their results were statistically significant due to low response rates and the district's inability to fully implement the random assignment. The study outcomes included more clarity for principals about the expectations related to evaluation and more specific feedback as part of the evaluation process. In addition, those assigned to the new system seemed to place greater emphasis on communicating a school vision and technology than those evaluated under the old system. Finally, the authors reported that principals evaluated

under the new system seemed to receive more support for improving their practice (Kimball, Milanowski, & McKinney, 2007).

In sum, despite increased interest in principal evaluation and principal leadership on the part of policymakers and researchers, few studies have investigated the relationship between evaluation and learning-centered leadership behaviors. The studies by Smith et al. (2008) and Kimball et al. (2007) represent promising initial efforts to consider the relationship between aspects of evaluation and principal outcomes. But they provide little insight into whether certain evaluation practices are likely to be associated with ambitious goals for student learning, demanding curricula, and/or support for and evaluation of instructional quality. The study presented here directly addresses this gap in the literature.

THEORETICAL FRAMEWORK

The research design for this study was based on a theoretical framework that articulates possible relationships between principal evaluation and learning-centered (LC) leadership (see Figure 1 for a visual representation of the framework). In this section, we first delineate the core principal behaviors associated with this type of leadership (Murphy et al., 2007). Then we draw on the framework to hypothesize ways that the purposes and foci of principal evaluation systems, and the leadership activities they assess, may be related to such leadership (Portin, Feldman, & Knapp, 2006). As we make evident in this section, our framework is based on an existing conception of effective school leadership (Murphy et al., 2007) that is closely connected to empirical research on principals.

Drawing on work by Murphy et al. (2007), our framework posits that LC leadership features several core behaviors. One core leadership behavior is to establish rigorous goals for student learning (Leithwood & Janzi, 2005). School leaders can increase teachers' understanding of and support for such goals by translating them into grade-specific objectives and delineating each teacher's role in helping students to meet them (Leithwood, 1994; Hallinger & Heck, 1996). They can also promote instructional quality and student learning by articulating a clear mission and building support for it among school staff (Hallinger, Bickman, & Davis 1996; Snyder & Ebmeier, 1992). A second central behavior is to coordinate curriculum in ways that ensure that all students have access to demanding academic content in the main content areas. School leaders can promote equitable access to ambitious content by helping teachers translate the school's academic goals into common curricular objectives and classroom activities (Murphy et al., 2007; Newmann, 1997).

Third, LC leadership includes promoting and enforcing high academic standards, or quality instruction, which is defined as teaching practices that

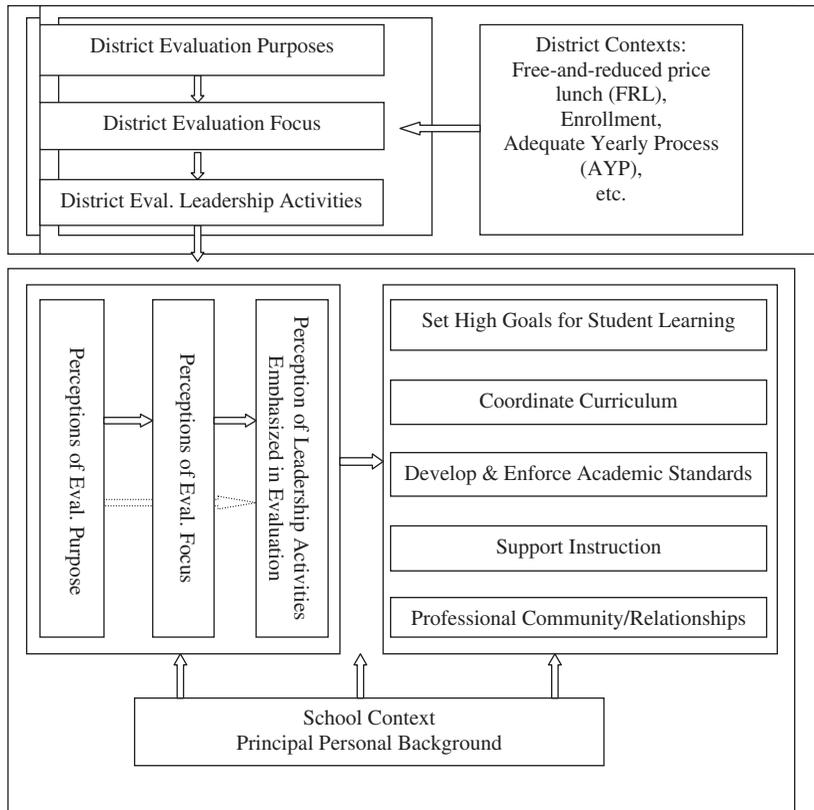


FIGURE 1 How district evaluation policy influences principal leadership practices.

maximize student engagement and learning. Principals can monitor the quality of instruction through classroom observations, postobservation conferences, and other direct contact with teachers (Burch & Spillane, 2003; Coburn, 2001; Stein & D'Amico, 2002; Youngs & King, 2002). This form of leadership also involves supporting instruction by using student and teacher data to provide feedback to teachers and to guide decision making (Heck, 1992; Leithwood & Jantzi, 2005). Finally, LC leadership includes analysis of school organizational conditions by principals and nurturing productive work relationships among teachers and between themselves and teachers. In particular, such leadership involves promoting professional community among teachers and relational trust between teachers and school administrators and between school staff and students and their families (Bryk, Camburn, & Louis, 1999; Bryk & Schneider, 2002; Hart, 1992; Louis, Marks, & Kruse, 1996).²

While increasing agreement is evident among researchers on how to define LC leadership, it is less clear from the literature how, if at all, principal evaluation is related to such leadership. Our framework indicates that

principal evaluation systems can be categorized based on their *purposes*; the principal characteristics, behaviors, and outcomes on which they *focus*; and the *leadership activities* that they address (Portin, Feldman, & Knapp, 2006). When principal evaluation is designed to promote school restructuring, inform principals' professional development, and/or hold leaders accountable for student performance, the framework suggests that evaluation practices will result in LC leadership (Murphy et al., 2007; Portin, Feldman, & Knapp, 2006).

As noted, principal evaluation practices in the United States have traditionally concentrated on school leaders' personal traits and characteristics, but recently districts and states have begun to focus more on their knowledge and skills, their leadership behaviors, the professional relationships they establish, and the outcomes of their work. Our framework posits that principal evaluation is likely to be related to LC leadership when it addresses leaders' knowledge and skills related to instruction, their leadership behaviors, and their relationships with teachers, parents, and others. Further, evaluation is also expected to promote such leadership when it considers principals' impact on school climate and student performance (Murphy et al., 2007; Portin, Feldman, & Knapp, 2006).

In terms of leadership activities, the framework proposes that principal evaluation that focuses on principals' ability to set instructional goals, design curriculum, evaluate teachers, and arrange teacher professional development is expected to result in LC leadership. Further, evaluation is also expected to promote such leadership when it addresses leaders' efforts to monitor instruction and communicate with staff about student performance (Murphy et al., 2007; Portin, Feldman, & Knapp, 2006).

SAMPLE

Data for the study came from two sources. Data on district enrollment size in 2006–2007 and the percentage of students in each district who were eligible for free or reduced-price lunch that same year were obtained from the Center for Education Performance and Information (CEPI) at the Michigan Department of Education. All other data used in the analyses came from a series of surveys administered by the researchers, including district administrator surveys, principal surveys, and teacher surveys.

In spring 2008, we administered surveys to 19 administrators in 13 Michigan school districts who had been responsible for the evaluation of elementary and/or secondary school principals in their districts in 2006–2007. Across these districts, we administered surveys to 138 principals who had served in 2006–2007 as principals in their buildings (and were continuing to serve in this role in 2007–2008). In each school, we administered surveys to two teachers who had worked in the school in 2006–2007 (and were

continuing to teach there in 2007–2008). The response rates for the district administrator, principal, and teacher surveys were 63.16 percent, 61.69 percent, and 60.96 percent, respectively.³

We initially sought districts that had more than eight K–12 public schools in 2006–2007 in order to increase the variability of data on both levels of the Hierarchical Linear Model (HLM), specifically, the district level and the individual principal level. As shown in Table 1, the final sample included 49 elementary schools, 17 middle schools, and 19 high schools. Moreover, as shown in Table 2, the participating districts varied in enrollment size and the percentage of students who were eligible for free or reduced lunch.

We administered electronic and hard copy surveys to district administrators and principals in spring, 2008. To ensure high response rates, a five-contact approach was employed (Dillman, 2007). This method involved sending each eligible district administrator and principal (a) an initial letter introducing the study; (b) a cover letter, a consent form, a link to the electronic survey, and a token incentive; (c) a postcard reminding eligible participants about the survey/thanking them for their participation; (d) a second mailing of the cover letter, consent form, and electronic survey link; and (e) and a final contact (hard copy survey with return envelope) to those who had not yet responded. The five contacts (a through e) took place one week apart from each other between March and May 2008. At each step (i.e., after each contact), about 20 percent of the (remaining) sample completed the survey.

The district administrator survey had several sections. The first section addressed the intentions of the district's principal evaluation policy, the second section included items about participants' perceptions of the effects of the evaluation system, and the third section asked about their personal and professional background. The principal survey also had three sections. The first section inquired into participants' perceptions of the district's principal evaluation system (the items in this section were identical to those in the first section of the district administrator survey). We used these items to assess policy outcomes; i.e., whether principals' perceptions of the district evaluation system were consistent with district intentions. The second section of the principal survey focused on principal leadership behaviors, while the third section asked about the school and the principal's personal and professional background.

The teacher surveys also included three sections. The first section featured items on the participants' instructional practices and their perceptions of school climate, while the second section inquired into teachers' perceptions of the principal's leadership at their school. (The items in the second section of the teacher survey were identical to those in the second section of the principal survey.) Finally, the third section included items on the teachers' personal and professional background.

TABLE 1 Number and Percentage of Schools in Each District by Educational Levels.

District ID	School levels	Number of schools in each district	Percentage of schools in each district (%)
1	Elementary	4	66.7
	Middle	1	16.7
	High	1	16.7
	Total	6	100.0
2	Elementary	5	71.4
	Middle	1	14.3
	High	1	14.3
	Total	7	100.0
3	Elementary	2	66.7
	Middle	1	33.3
	Total	3	100.0
4	Elementary	4	40.0
	Middle	3	30.0
	High	3	30.0
	Total	10	100.0
5	Elementary	3	60.0
	High	2	40.0
	Total	5	100.0
6	Elementary	4	50.0
	Middle	4	50.0
	Total	8	100.0
7	Elementary	5	50.0
	Middle	2	20.0
	High	3	30.0
	Total	10	100.0
8	Elementary	2	66.7
	High	1	33.3
	Total	3	100.0
9	Elementary	2	40.0
	Middle	1	20.0
	High	2	40.0
	Total	5	100.0
10	Elementary	3	50.0
	Middle	1	16.7
	High	2	33.3
	Total	6	100.0
11	Elementary	4	66.7
	Middle	1	16.7
	High	1	16.7
	Total	6	100.0
12	Elementary	6	75.0
	Middle	1	12.5
	High	1	12.5
	Total	8	100.0
13	Elementary	5	62.5
	Middle	1	12.5
	High	2	25.0
	Total	8	100.0

TABLE 2 District Enrollment Size and Percentage of Students Who Were Eligible for Free or Reduced Price Lunch in 2006–2007.

District IDs	Enrollment	Percent of FRP lunch (%)
1	16,656	15.90
2	6,944	66.75
3	4,408	11.77
4	12,090	10.96
5	8,835	4.93
6	10,569	16.70
7	17,088	8.73
8	18,696	8.19
9	11,187	42.29
10	10,542	28.74
11	15,654	10.89
12	15,304	28.51
13	13,541	45.28

MEASURES

We built a Hierarchical Linear Model (HLM) with two-level and three-level applications to examine the impact of district principal evaluation policy on principals' leadership behaviors. In this section, we present the dependent measures and the predictors used in the model. Each of the dependent variables was associated with a key aspect of learning-centered leadership.

Dependent Variable 1: Setting High Goals for Student Learning

Five items in the principal survey measured principals' perceptions of the extent to which they established and maintained rigorous goals for student learning. The items were as follows: "Develop goals that seek improvement over current levels of academic performance," "Frame the school's academic goals in terms of staff responsibilities for meeting them," "Develop goals that are easily translated into classroom objectives by teachers," "Refer to the school's academic goals in informal settings with teachers (e.g., coffee hours)," and "Discuss the school's academic goals in formal settings with teachers (e.g., faculty meetings)." For these and the other items associated with the dependent variables (described below), principals rated their leadership according to the following 4-point scale: "not at all," "some extent," "moderate extent," and "great extent." These five items were aggregated into one composite dependent variable by taking the mean based on factor analysis⁴ ($\alpha = 0.85$).

Dependent Variable 2: Coordinating Curriculum

Principals' perceptions of their efforts to enact and coordinate a demanding academic curriculum were represented by the mean rating of five items ($\alpha = 0.82$). The items included "Make clear who is responsible for coordinating the curriculum across grade levels (e.g., assistant principal, or a teacher)," "Ensure that the school's academic goals are translated into common curricular objectives and classroom activities," "Draw on the results of schoolwide testing when making curricular decisions," "Assess the overlap between the school's curricular objectives and the achievement test(s) used for program evaluation," and "Participate actively in reviewing or selecting curricular materials."

Dependent Variable 3: Developing and Enforcing Academic Standards

The measure of how principals developed and enforced academic standards to promote student learning was the mean of two components: "Make known what is expected of students at different grade levels" and "Support teachers when they enforce academic policies" ($\alpha = 0.80$).

Dependent Variable 4: Supporting Instruction

Principals' perceptions of their efforts to support and monitor instruction were represented by the mean of six items, including "Point out specific strengths in teacher instructional practices in post observation conferences," "Note specific weaknesses of the teacher's instructional practices in written evaluations," "Distribute journal articles to teachers on a regular basis," "Arrange for outside speakers to make presentations on instruction at faculty meetings," "Provide time to discuss instructional issues with teachers," and "Develop or find the appropriate instructional program(s) for students whose test results indicate a need" ($\alpha = 0.83$).

Independent Measures

As we mentioned before, both the district administrator survey and the principal survey included items concerning principal evaluation policy. We used the same items to measure principals' perceptions of district evaluation policy (policy outcomes) and district administrators' perceptions of the policy (policy intentions).

Predictor 1: Principal Perceptions of the Purpose of Evaluation

As noted, based on previous research, we expected that when the purpose of principal evaluation was to promote school restructuring, support school leader professional development, and/or hold leaders accountable for student

learning, evaluation was likely to lead to learning-centered leadership behaviors. Items in the principal survey asked whether principal evaluation addressed one or more of the following purposes: (a) to provide information for use in making decisions about hiring, principal promotion, (re)assignment, and contract renewal within the district; (b) to provide evidence for use in determining merit salary increases or sanctions for principals; (c) to promote the professional development of principals; (d) to facilitate school restructuring; and (e) to hold principals accountable for student achievement. For these and the items associated with the other three predictor variables (described below), principals rated their district's principal evaluation system according to the following 4-point scale: "not at all," "some extent," "moderate extent," and "great extent." Factor analysis indicated that three of these items—c, d, and e—were one factor ($\alpha = 0.76$). We then aggregated them into one composite variable by the mean.

Predictor 2: Principal Perceptions of the Focus of Evaluation

We wanted to examine whether principal evaluation could help school leaders to develop learning-centered leadership behaviors if it emphasized their (a) knowledge and skills, (b) behaviors, and (c) professional relationships, as opposed to their individual traits, personality, or dispositions. Further, we also wanted to consider whether a focus in evaluation on (d) school organizational impacts on school climate and student achievement would be related to learning-centered leadership. One composite variable was created by taking the sum of six items that addressed a, b, c, and d ($\alpha = 0.85$).

Predictor 3: Principal Perceptions of the Leadership Activities Evaluated by the District

We also wanted to investigate whether principal evaluation that addressed instructional leadership activities had an effect on learning-centered leadership. Four items inquired into principals' perceptions of the extent to which district evaluation focused on noninstructional leadership activities, including communication with community organizations, faculty meetings, school assemblies, and extracurricular activities. On the other hand, six items asked about school leaders' perceptions regarding the extent to which district evaluation covered instructional leadership activities, including school goal setting, curriculum design, teacher evaluation, providing professional development for teachers, monitoring student learning during school time, and focusing on meeting Adequate Yearly Progress. One composite variable was created by taking the sum of the six items that addressed instructional leadership activities ($\alpha = 0.89$).

Two Covariates at the Principal Level: Principal's Working Experience as a Principal and Principal's Working Experience as a Full-Time Teacher in K–12 Schools

Principals could rate their working experiences as a principal and as a full-time K–12 teacher according to the following 5-point scale: less than 1 year, 1–3 years, 4–8 years, 9–12 years, and more than 13 years.

Five Covariates at the District Level

We included five district-level covariates in some models, including district evaluation purposes, evaluation foci, evaluated leadership activities, enrollment size, and the percentage of students who were eligible for free or reduced lunch.

To measure district policy intentions, we aggregated three covariates from the district administrator surveys by using the same items as those used to generate the predictors from the principal surveys: evaluation purposes, evaluation foci, and evaluated leadership activities. Moreover, data on district enrollment size and the percentage of students eligible for free or reduced lunch were generated from the 2006–2007 CEPI data.

Missing Data

We had a small amount of missing data (from 4.7 percent to 5.9 percent) for each of the four dependent variables: setting high goals for student learning, coordinating curriculum, developing and enforcing academic standards, and supporting instruction. For the five independent variables and the covariates at the individual level, the percentage of missing data ranged from 5.2 to 8.3 percent. The district-level variables of “enrollment size” and “students eligible for free/reduced lunch” did not contain any missing data, while we had a relatively large proportion of missing values (16.1 percent) for each of the following variables: “district evaluation purpose,” “district evaluation focus,” and “district leadership activities emphasized in district evaluation.”

Contextual replacement methods (Frank et al., 2008) were used to account for the missing data. Basically, when values were missing from a principal survey with regard to principal leadership behaviors (level-1), we used the average of teachers' ratings of principal leadership in that school to replace the missing values. Similarly, when values were missing from a district administrator survey with regard to principal evaluation policy, we used the average of principals' ratings for evaluation policy in that district to replace the missing values. In comparison to other statistical imputation methods for replacing missing values, contextual replacement methods were most appropriate for this study for a few reasons. First, policy and

leadership were undertaken within local social and political contexts in this study (Spillane, Reiser, & Reimer, 2002; Weatherley & Lipsky, 1977). Thus, it made more sense to use the average of teachers' responses in a specific school to replace missing data for the leadership behavior of that school's principal than to use other principals' behaviors from different school contexts to estimate the principal's behavior. Similarly, at the district level, it made more sense to use the average of principals' responses about district evaluation policy than to use data on evaluation policy from other district contexts.

Second, as discussed previously, there was considerable overlap between the items on the teacher and principal surveys and between the items on the principal and superintendent surveys. As a result, we had sufficient data from different parties (e.g., teachers and principals) to enable us to use replacement methods. Third, in a study (such as the one reported here) with a relatively small sample size, replacement methods that rely on a large sample size might not be suitable. Finally, we did not replace missing values for the variables of "working experience as a principal" or "working experience as a full-time teacher" because we did not have other information on these two variables.

MODELS

Due to the multilevel nature of the research design (repeated observations nested with principals, principals nested in districts), we built Hierarchical Linear Models (HLM)⁵ to investigate the effect of evaluation policy and practices on principal leadership (Raudenbush & Bryk, 2002). First, we built HLM models (HLM at its two-level application) for each of the four dependent variables (i.e., setting high goals for student learning, coordinating curriculum, developing and enforcing academic standards, and supporting instruction and learning) with same predictors and same covariates. These models were intended to help us understand the impact of evaluation policy on leadership in different areas. Second, because of the strong correlation among dependent variables, we built a two-level multivariate model (HMLM) to estimate how policy affected different leadership behaviors simultaneously. We then added covariate variables into this two-level multivariate model to develop a three-level HMLM (HMLM2) model, which allowed us to examine the interaction between evaluation policy intentions and outcomes.

Before finalizing the first set of models described above, we formulated four unconditional models to examine variance components associated with each dependent variable (see Table 3). In general, for these four models, approximately 90 percent of the variance was located at level-1 (individual principal level), and about 10 percent of the variance at level-2 (district

TABLE 3 Variance Components of Four HLM Models.

Models	Level-1 variance	Level-2 variance	Total variance	Percent of level-1 variance (%)	Percent of level-2 variance (%)	Chi-square level-1 vs. level-2
Model 1	0.280	0.043	0.323	86.54	13.46	24.230 (0.019)**
Model 2	0.352	0.036	0.388	90.72	9.28	20.093 (0.065)*
Model 3	0.284	0.008	0.292	97.29	2.71	13.800 (0.313)
Model 4	0.234	0.031	0.264	88.44	11.56	22.154 (0.036)**

Notes: The dependent variable of model 1 is “setting high goals for student learning.” Dependent Variable in model 2 is “coordinating curriculum,” in model 3 “developing and enforcing academic standards,” and in model 4 “supporting instruction.”

P-values of chi-square statistics are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

level). In the model with the dependent variable “developing and enforcing academic standards,” the proportion of variance at level-2 was much smaller relative to that at level-1. But the variance components at level were significant in the other three models.

We then tested models by adding predictors at level-1 and covariates at level-2. As shown in Table 4, a significant correlation was evident among three predictors: principals’ perceptions of evaluation purpose, perceptions of evaluation focus, and perceptions of the leadership activities emphasized by evaluation system. In view of the strong correlation among these three predictors, we decided to add them separately into the model to avoid collinearity, along with two covariates (working experience as a principal and working experience as a teacher) at level-1, and two covariates at level-2 (district enrollment size and the percentage of students eligible for free or reduced lunch). The general model can be specified as follows:

Level-1 Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Perceived Eval Purposes}_{ij}) + \beta_{2j}(\text{Perceived Eval Foci}_{ij}) \\ + \beta_{3j}(\text{Perceived Eval Leadership Activities}_{ij}) + \beta_{4j}(\text{Exp as Principal}_{ij}) \\ + \beta_{5j}(\text{Exp as Teacher}_{ij}) + \varepsilon_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Enroll}_j) + \gamma_{02}(\text{FRL}_j) + v_{0j}$$

$$\beta_{1j} = \gamma_{01}$$

$$\beta_{2j} = \gamma_{02}$$

$$\beta_{3j} = \gamma_{03}$$

$$\beta_{4j} = \gamma_{04}$$

$$\beta_{5j} = \gamma_{05}$$

TABLE 4 Pearson Correlation Coefficients Among Independent Variables.

	Principal perception of evaluation purpose	Principal perception of evaluation focus	Principal perception of leadership acts. emphasized in evaluation
Principal Perception of Evaluation Purpose	1	0.378 (0.0003)***	0.505 (< 0.0001)***
Principal Perception of Evaluation Focus	0.378 (0.0003)***	1	0.484 (< 0.0001)***
Principal Perception of Leadership Acts. Emphasized in Evaluation	0.505 (< 0.0001)***	0.484 (< 0.0001)***	1

Notes: P-values are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

Y_{ij} : Principal i 's leadership behaviors in district j ; behaviors can include dependent variable 1 (setting high goals for student learning), dependent variable 2 (coordinating curriculum), dependent variable 3 (developing and enforcing academic standards), and dependent variable 4 (supporting instruction)

β_{0j} : The intercept

β_{1j} : The effect of perceptions of principal i in district j of evaluation purpose

β_{2j} : The effect of perceptions of principal i in district j of evaluation focus

β_{3j} : The effect of perceptions of principal i in district j of leadership activities emphasized in evaluation

β_{4j} : The effect of years of experience working as a principal for principal i in district j

β_{5j} : The effect of years of experience working as a full-time teacher for principal i in district j

ε_{ij} : The error term at level-1

γ_{01} : The effect of enrollment size in district j

γ_{02} : The effect of percentage of students eligible for free or reduced lunch in district j

ν_{0j} : The error term at level-2

After examining dependent variables individually, we built a two-level HMLM model to examine how predictors and covariates affected the four dependent variables simultaneously. It is important to account for the four dependent variables and the correlations among them in the two-level HMLM model because the four dependent variables are highly correlated with each other, as shown in Table 5. In other words, we treated these four dependent variables as a single construct, which can be referred to as learning-centered leadership.

In addition, we compared three different models (an unrestricted model, a homogeneous level-1 variance model, and a heterogeneous level-1 variance

TABLE 5 Pearson Correlation Coefficients among Dependent Variables.

	Setting high goals for student learning	Coordinating curriculum	Developing / enforcing academic standards	Supporting instruction
Setting High Goals for Student Learning	1	0.685 (< 0.0001)***	0.577 (< 0.0001)***	0.535 (< 0.0001)***
Coordinating Curriculum	0.685 (< 0.0001)***	1	0.498 (< 0.0001)***	0.616 (< 0.0001)***
Developing /Enforcing Academic Standards	0.577 (< 0.0001)***	0.498 (< 0.0001)***	1	0.493 (< 0.0001)***
Supporting Instruction	0.535 (< 0.0001)***	0.616 (< 0.0001)***	0.493 (< 0.0001)***	1

Notes: P-values are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

model) and concluded that the unrestricted model fit the data better than the homogeneous and heterogeneous models. This is because χ^2 statistics showed that we had to reject the hypotheses and accept the alternative hypothesis that the unrestricted model was a better fit than a) the homogeneous model ($\chi^2 = 32.275$ and $p\text{-value} \leq 0.0005$) and b) the heterogeneous model ($\chi^2 = 23.145$ and $p\text{-value} \leq 0.001$). The model can be simplified as follows:

Level-1 Model

$$Y_{mi} = (\text{IND1}_{mi})(\text{Set Goals}_{1i})^* + (\text{IND2}_{mi})(\text{Coordinate Curriculum}_{2i})^* \\ + (\text{IND3}_{mi})(\text{Academic Standards}_{3i})^* + (\text{IND4}_{mi})(\text{Support Instruction}_{4i})^*$$

$$Y^*_{mi} = \pi_{0i} + \varepsilon_{mi}$$

Level-2 Model

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Perceived Eval Purposes}_j) + \beta_{02}(\text{Perceived Eval Foci}_j) \\ + \beta_{03}(\text{Perceived Eval Leadership Activities}_j) + \beta_{04}(\text{Exp as Principal}_j) \\ + \beta_{05}(\text{Exp as Teacher}_j)$$

Y_{mi} : The dependent variables in this study, m , can be 1st, 2nd, 3rd, and 4th

IND1: Indicator of the 1st dependent variable

IND2: Indicator of the 2nd dependent variable

IND3: Indicator of the 3rd dependent variable

IND4: Indicator of the 4th dependent variable

π_{0i} : The intercept at level-1

ε_{mi} : The error term at level-1

- β_{00} : The intercept at level-2
 β_{01} : The effect of perceptions of principal j of evaluation purpose
 β_{02} : The effect of perceptions of principal j of evaluation focus
 β_{03} : The effect of perceptions of principal j of leadership activities emphasized in evaluation
 β_{04} : The effect of experiences as a principal by principal j
 β_{05} : The effect of experiences as a full-time teacher by principal j

Finally, we added a third level to the previous two-level HLM by including the covariates at the district level. This model enabled us to examine the interaction between evaluation policy intentions (i.e., district administrators' perceptions of evaluation policy) and policy outcomes (as perceived by principals). We also compared three different models (an unrestricted model, a homogeneous level-1 variance model, and a heterogeneous level-1 variance model) and concluded that the unrestricted model better fit our data than homogeneous and heterogeneous models. This was due to the fact that χ^2 statistics provided support for rejecting the hypotheses and accepting the alternative hypothesis that the unrestricted model was a better fit than a) the homogeneous model ($\chi^2 = 31.66$, p-value ≤ 0.0005) or b) the heterogeneous model ($\chi^2 = 24.41$, p-value ≤ 0.0005) as detailed in Table 4. The model can be simplified as:

Level-1 Model

$$Y_{mij} = (\text{IND1}_{mij}) (\text{Set Goals1}_{ij})^* + (\text{IND2}_{mij}) (\text{Coordinate Curriculum2}_{ij})^* \\ + (\text{IND3}_{mij}) (\text{Academic Standards3}_{ij})^* + (\text{IND4}_{mij}) (\text{Support Instruction4}_{ij})^*$$

$$Y^*_{mij} = \pi_{0ij} + \varepsilon_{mij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{Perceived Eval Purposes}_{ij}) + \beta_{02j}(\text{Perceived Eval Foci}_{ij}) \\ + \beta_{03j}(\text{Perceived Eval Leadership Activities}_{ij}) + \beta_{04j}(\text{Exp as Principal}_{ij}) \\ + \beta_{05j}(\text{Exp as Teacher}_{ij})$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\text{Enroll}_j) + \gamma_{002}(\text{FRL}_j) + v_j$$

$$\beta_{01j} = \gamma_{010} + \gamma_{011}(\text{District Eval Purposes}_j)$$

$$\beta_{02j} = \gamma_{020} + \gamma_{021}(\text{District Eval Contents}_j)$$

$$\beta_{03j} = \gamma_{030} + \gamma_{031}(\text{District Eval Leadership Activities}_j)$$

$$\beta_{04j} = \gamma_{040}$$

$$\beta_{05j} = \gamma_{050}$$

- γ_{011} : The impact of district intended evaluation purpose on principals' perceptions of evaluation purpose
- γ_{021} : The impact of district intended evaluation focus on principals' perceptions of evaluation focus
- γ_{031} : The impact of leadership activities intended by district to be emphasized in evaluation on principals' perceptions of leadership activities emphasized in evaluation

RESULTS

Descriptive Results

Table 6 shows the descriptive statistics for all variables included in the HLM models. The means of the four leadership behaviors were between 3 and 4 on a 4-point scale, indicating that the sampled principals enacted LC leadership activities from a "moderate extent" to a "great extent." These principals perceived that district evaluation did not strongly emphasize principal professional development or student learning. In addition, they did not perceive district evaluation as being highly focused on leadership outcomes or

TABLE 6 Descriptive Statistics of All Variables.

Level/Variable	Number	Mean	Standard deviation
<i>Individual Level</i>			
<i>Dependent Variables</i>			
Setting High Goals for Student Learning	85	3.47	0.57
Coordinating Curriculum	85	3.27	0.62
Developing and Enforcing Academic Standards	85	3.72	0.54
Supporting Instruction	85	3.26	0.51
HMLM and HMLM2 (Including Four Dependent Variables)	340	3.43	0.59
<i>Independent Variables and Covariates</i>			
Principal Perception of Evaluation Purpose	85	2.72	1.06
Principal Perception of Evaluation Focus	85	4.42	2.02
Principal Perception of Leadership Acts. Emphasized in Evaluation	85	2.76	0.99
Working Experience as a Principal	85	3.41	1.31
Working Experience as a Full-time teacher	85	3.84	1.23
<i>District Level</i>			
Enrollment Size	13	12424.15	4230.52
Students Eligible for Free/Reduced Price Lunch	13	0.23	0.18
District Evaluation Purpose	13	3.20	0.74
District Evaluation Focus	13	22.18	10.16
District Leadership Acts. Emphasized in District Evaluation	13	3.30	0.50

LC leadership activities. Finally, it is clear that a diverse group of principals was sampled in terms of their previous working experiences as principals and full-time teachers because standard deviations for these two variables were relatively large.

At the district level, the sampled districts varied in enrollment size and percentage of students who were eligible for free or reduced lunch. However, in terms of focusing on principal professional development and LC leadership behaviors, district administrators generally reported higher ratings for evaluation purpose and focus than principals did.

Results of Individual Models

As reported in Tables 7–10, we used four separate models to predict the likelihood of principals' responses regarding such LC leadership behaviors as setting goals for student learning, coordinating curriculum, developing and enforcing academic standards, and supporting student learning. These four models produced similar results; in each case, five important variables

TABLE 7 HLM Model with Dependent Variable: Setting High Goals for Student Learning.

Level/Variable	Estimates		
	Model-1	Model-2	Model-3
<i>Principal level (level-1, N = 85)</i>			
Principal Perception of Evaluation Purpose	0.121 (0.057)**	—	—
Principal Perception of Evaluation Focus	—	0.073 (0.031)**	—
Principal Perception of Leadership Acts. Emphasized in Evaluation	—	—	0.155 (0.059)***
Working Experience as a Principal	0.007 (0.053)	-0.016 (0.055)	-0.004 (0.053)
Working Experience as a Full-time Teacher	0.114 (0.054)**	0.123 (0.052)**	0.112 (0.052)**
<i>District Level (level 2, N = 13)</i>			
Enrollment Size	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Percentage of Students Eligible for Free/Reduced Price Lunch	0.884 (0.470)*	1.209 (0.459)**	1.057 (0.457)*

Notes: The dependent variable is the same in each of these three models: setting high goals for student learning. The independent variables in model 1 do not include "principal perception of evaluation focus" and "principal perception of leadership activities (acts.) emphasized in evaluation." The independent variables in model 2 do not include "principal perception of evaluation purposes" and "principal perception of leadership acts. emphasized in evaluation." The independent variables in model 3 do not include "principal perception of evaluation purposes" and "principal perception of evaluation focus."

Standard errors are reported in parentheses. *p ≤ 0.1; **p ≤ 0.05; ***p ≤ 0.001.

TABLE 8 HLM Model with Dependent Variable: Coordinating Curriculum.

Level/Variable	Estimates		
	Model-1	Model-2	Model-3
<i>Principal Level (Level 1, N = 85)</i>			
Principal Perception of Evaluation Purpose	0.177 (0.062)**	—	—
Principal Perception of Evaluation Focus	—	0.097 (0.035)**	—
Principal Perception of Leadership Act. Emphasized in Evaluation	—	—	0.265 (0.062)***
Working Experience as a Principal	0.008 (0.052)	-0.118 (0.060)*	-0.109 (0.055)*
Working Experience as a Full-time Teacher	0.052 (0.052)	0.121 (0.059)**	0.092 (0.057)
<i>District Level (Level 2, N = 13)</i>			
Enrollment Size	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Percentage of Students Eligible for Free/ Reduced Price Lunch	1.021 (0.386)**	0.097 (0.035)	1.326 (0.356)**

Notes: The dependent variable is the same in each of these three models: coordinating curriculum. The independent variables in model-1 do not include “principal perception of evaluation focus” and “principal perception of leadership activities (acts.) emphasized in evaluation.” The independent variables in model-2 do not include “principal perception of evaluation purposes” and “principal perception of leadership acts. emphasized in evaluation.” The independent variables in model-3 do not include “principal perception of evaluation purposes” and “principal perception of evaluation focus.”

Standard errors are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

consistently predicted the extent to which principals engaged in each of these four types of LC leadership behaviors. The five variables included principal’s perceptions of evaluation purposes, principal’s perceptions of evaluation focus, principal’s perceptions of leadership activities emphasized in evaluation, principal’s working experiences as a full-time teacher, and percent of students eligible for free or reduced-price lunch district wide. Since the results were similar in these four models, we use model-1 here (with the dependent variable of setting high goals for student learning) to illustrate how to interpret the results. It is important to note, though, that these variables varied in degree of significance and for each, their degree of significance differed slightly across models.

As shown in Table 7, by holding other variables in the model constant, we found that principal’s perception (regarding evaluation purpose) that district evaluation aims to promote principals’ professional development, facilitate school restructuring, and hold principals accountable for student learning significantly predicted the likelihood that the principal would set high goals for learning. Specifically, with a one-unit increase in the principal’s perception that the purposes of district evaluation were to promote principal professional development, strengthen leaders’ impact on school

TABLE 9 HLM Model with Dependent Variable: Developing and Enforcing Academic Standards.

Level/Variable	Estimates		
	Model-1	Model-2	Model-3
<i>Principal Level (Level 1, N = 85)</i>			
Principal Perception of Evaluation Purposes	0.157 (0.056)**	—	—
Principal Perception of Evaluation Focus	—	0.091 (0.030)**	—
Principal Perception of Leadership Acts. Emphasized in Evaluation	—	—	0.153 (0.059)**
Working Experience as a Principal	-0.028 (0.051)	-0.061 (0.052)	-0.038 (0.052)
Working Experience as a Full-time Teacher	0.131 (0.053)**	0.146 (0.051)**	0.138 (0.053)**
<i>District Level (Level 2, N=13)</i>			
Enrollment Size	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Percentage of Students Eligible for Free/ Reduced Price Lunch	0.385 (0.343)	0.814 (0.336)**	0.630 (0.334)*

Notes: The dependent variable is the same in each of these three models: developing and enforcing academic standards. The independent variables in model-1 do not include “principal perception of evaluation focus” and “principal perception of leadership activities (acts.) emphasized in evaluation.” The independent variables in model-2 do not include “principal perception of evaluation purposes” and “principal perception of leadership acts. emphasized in evaluation.” The independent variables in model-3 do not include “principal perception of evaluation purposes” and “principal perception of evaluation focus.”

Standard errors are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

restructuring, and hold them accountable for learning, the likelihood that the principal would focus on setting high learning goals for student learning increased by 0.12 units, with standard error of 0.06 and p-value less than 0.05.

Similarly, by holding other variables constant, with a one-unit increase in principal’s perception that their district evaluated principals based on their knowledge and skills, behaviors, and professional relationships (as opposed to their personal traits), the probability that the principal would focus on setting high goals for student learning increased by 0.07 units, with standard error of 0.03 and p-value less than 0.05. Finally, with a one-unit increase in the principal’s perception that evaluation stressed instructional leadership activities (i.e., setting school goals and designing curriculum, monitoring student learning, and evaluating teachers and arranging professional development for them), the principal was 0.15 units more likely to focus on setting high goals for learning, with standard error of 0.06 and p-value less than 0.001.

Principal experience working as a full-time teacher in K–12 was also a strong and consistent predictor of the dependent variable. With more

TABLE 10 HLM Model with Dependent Variable: Supporting Instruction.

Level/Variable	Estimates		
	Model-1	Model-2	Model-3
<i>Principal Level (Level 1, N = 85)</i>			
Principal Perception of Evaluation Purpose	0.133 (0.054)**	—	—
Principal Perception of Evaluation Focus	—	0.045 (0.029)	—
Principal Perception of Leadership Acts. Emphasized in Evaluation	—	—	0.150 (0.055)**
Working Experience as a Principal	0.045 (0.050)	0.038 (0.052)	0.041 (0.049)
Working Experience as a Full-time Teacher	0.055 (0.051)	0.095 (0.045)**	0.079 (0.044)*
<i>District Level (Level 2, N=13)</i>			
Enrollment Size	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Percentage of Students Eligible for Free/ Reduced Price Lunch	0.517 (0.368)	0.806 (0.366)*	0.751 (0.322)*

Notes: The dependent variable is the same in each of these three models: supporting instruction. The independent variables in model-1 do not include “principal perception of evaluation focus” and “principal perception of leadership activities (acts.) emphasized in evaluation.” The independent variables in model-2 do not include “principal perception of evaluation purposes” and “principal perception of leadership acts. emphasized in evaluation.” The independent variables in model-3 do not include “principal perception of evaluation purposes” and “principal perception of evaluation focus.”

Standard errors are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

experience as a full-time teacher, the principal might have been more likely to frame high academic goals and effectively communicate them to teachers and students. However, working experience as a principal did not significantly affect the likelihood that the school leader would focus on setting high goals for student learning.

At the district level, district enrollment sizes did not seem to affect principal leadership behaviors on average in a given district. However, the higher the percentage of students who were eligible for free or reduced lunch, the higher the likelihood that principals set high goals for student learning.

HMLM Results

Because of the high and significant correlation among the four dependent variables, we built a hierarchical multivariate linear model (HMLM) with its two-level application to simultaneously model them. Results are shown in Table 11. Consistent with the results of the individual HLM models, principals' perceptions of district evaluation policy significantly predicted the

TABLE 11 HMLM with Four Dependent Variables.

Level/Variable	Estimates		
	Model-1	Model-2	Model-3
<i>Principal Level (Level 2, N = 85)</i>			
Principal Perception of Evaluation Purpose	0.156 (0.042)***	—	—
Principal Perception of Evaluation Focus	—	0.050 (0.024)**	—
Principal Perception of Leadership Acts. Emphasized in Evaluation	—	—	0.152 (0.046)***
Working Experience as a Principal	0.010 (0.037)	-0.009 (0.040)	-0.013 (0.038)
Working Experience as a Full-time Teacher	0.083 (0.041)**	0.116 (0.041)**	0.097 (0.041)**

Notes: The independent variables in model-1 do not include “principal perception of evaluation focus” and “principal perception of leadership activities (acts.) emphasized in evaluation.” The independent variables in model-2 do not include “principal perception of evaluation purposes” and “principal perception of leadership acts. emphasized in evaluation.” The independent variables in model-3 do not include “principal perception of evaluation purposes” and “principal perception of evaluation focus.”

Standard errors are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

extent to which they engaged in LC leadership behaviors. For instance, accounting for other variables in the model, with a one-unit increase in principal’s perception that the purpose of district evaluation was to promote principals’ professional development, facilitate school restructuring, and hold principals accountable for learning, a 0.15-unit increase was evident in the extent to which the principal would perform LC leadership behaviors (standard error = 0.04, p -value < 0.001). Similarly, the coefficient for principals’ perception of evaluation focus in predicting the degree of LC leadership was 0.05 (standard error = 0.02, p -value < 0.05) and the coefficient for principals’ perception of evaluated leadership activities in predicting the extent of LC leadership was 0.15 (standard error = 0.05, p -value < 0.001). Finally, principal’s working experience as a full-time teacher increased the probability that he or she would engage in LC leadership.

HMLM2 Results

The HMLM2 model was built to examine whether district policy intentions could affect policy outcomes, such as significantly affecting principals’ perceptions and leadership behaviors. We found no significant interactions between district evaluation policies (i.e., intentions) and principals’ perceptions of evaluation practices; the three across-level coefficients were almost equal to zero (see the last three rows in Table 12). In other words, district policy had little impact on principals’ perceptions of evaluation practices.

TABLE 12 HMLM2 with Interactions Between Policy Intentions and Policy Outcomes.

Level/Variable	Estimate		
	Model-1	Model-2	Model-3
<i>Principal Level (Level 2, N = 85)</i>			
Principal Perception of Evaluation Purpose	0.145 (0.063)**	—	—
Principal Perception of Evaluation Focus	—	0.114 (0.043)***	—
Principal Perception of Leadership Acts. Emphasized in Evaluation	—	—	0.173 (0.072)**
Working Experience as a Principal	-0.022 (0.057)	-0.058 (0.057)	-0.036 (0.056)
Working Experience as a Full-time Teacher	0.102 (0.057)*	0.131 (0.056)**	0.103 (0.057)*
<i>District Level (Level 3, N = 13)</i>			
Enrollment Size	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Percentage of Students Eligible for Free/ Reduced Price Lunch	0.831 (0.375)**	1.309 (0.380)**	0.962 (0.353)**
District Evaluation Purpose	-0.0247 (0.066)	—	—
District Evaluation Focus	—	-0.012 (0.006)	—
Leadership Activities Emphasized in District Evaluation	—	—	0.059 (0.133)

Notes: The independent variables in model-1 do not include “principal perception of evaluation focus” and “principal perception of leadership activities (acts.) emphasized in evaluation” at either the principal level or the district level. The independent variables in model-2 do not include “principal perception of evaluation purposes” and “principal perception of leadership acts. emphasized in evaluation” at either level. The independent variables in model-3 do not include “principal perception of evaluation purposes” and “principal perception of evaluation focus” at either level.

Standard errors are reported in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

The other predictors and covariates that were previously found to be significant (in the preceding models) continued to be important predictors in the HMLM2 model, after controlling for district policy intentions. These predictors and covariates included the three variables pertaining to perceptions of evaluation practices, the variable of working experience as a teacher (at the principal level), and percentage of students eligible for free or reduced lunch (at the district level).

DISCUSSION

As reported in the previous section and Tables 7–10, the findings from the individual models in this study suggest that the purpose and focus of principal evaluation in the 13 participating districts, and the leadership activities assessed, seemed related to key aspects of learning-centered (LC) leadership.

With regard to purpose, the results indicate that efforts by districts in the sample to use evaluation to guide school leaders' professional development (PD), to encourage school restructuring, and to hold leaders accountable for student learning were highly associated with LC leadership. These results build on other research that demonstrates the importance for districts of providing principal PD related to instructional leadership and ensuring that school leaders have the necessary knowledge and skills to translate broad learning goals into grade-level objectives and to monitor and support teachers' instructional practices (Coburn, 2001; Youngs & King, 2002).

The results also indicate an impact when principal PD was combined with accountability for student learning. By including accountability as a purpose of evaluation, the districts in the sample seemed to communicate to principals that their efforts to enact demanding curricula and monitor teaching practice would be assessed not only in their own right, but also with regard to their outcomes (i.e., their impacts on student learning). While current accountability policies in the United States have received much criticism, it may be that some combination of external and internal accountability systems can provide important guidance to school leaders and help ensure that principal PD leads to valued outcomes (Carnoy, Elmore, & Siskin, 2003; Newmann, King, & Rigdon, 1997).

In terms of the focus of principal evaluation, the findings suggest that concentrating on school leaders' knowledge, skills, and behaviors, as well as their professional relationships, was likely to be associated with LC leadership. This means districts may benefit from moving beyond the traditional emphasis in evaluation on principals' individual attributes, personality, and/or dispositions. In addition, our results indicate that the evaluation foci of (a) principals' knowledge and skills in enacting LC leadership (i.e., input), (b) their leadership activities (i.e., process), and (c) their impacts on school climate and student performance (i.e., outcomes), are closely related to each other. As one aspect of district evaluation policy, these foci could notably positively affect various aspects of LC leadership such as enacting ambitious learning goals, coordinating curricula, and promoting high quality instruction.

With regard to the leadership activities assessed, the results provide evidence that when activities related to instruction were emphasized by the districts in the sample, principals were likely to engage in LC leadership. In particular, school leaders were more likely to exhibit such leadership when district evaluation stressed school goal setting, curriculum design, teacher PD and evaluation, and monitoring student performance. This result is consistent with prior scholarship suggesting that evaluation oriented towards instructional leadership activities can strengthen principals' effects on teaching and learning. In addition, the magnitude of the coefficients for assessed leadership activities were the same or larger than those for evaluation purposes and foci. Although this study could not provide empirical data to explain such differences, one possible reason might be that compared with the evaluation purposes and

foci, the assessed leadership activities were more concrete and specific. Therefore, principals may have easily perceived connections between their daily activities and assessed leadership activities. Again, this finding suggests the potential gain for districts of moving beyond non-instructional leadership activities in evaluation to feature leadership activities that directly correspond to curriculum, teaching, and student learning (Portin, Feldman, & Knapp, 2006).

In several of the models presented in Tables 7–10, principals' prior work experience as a teacher was related to the outcomes of interest. In particular, this finding held up in the individual models for two aspects of LC leadership (setting high goals for student learning, and developing and enforcing academic standards). As discussed above, having more teaching experience may have helped school leaders devise rigorous academic goals and enforce high standards because they may have had greater knowledge of discrete learning objectives across grade levels and content areas. On the other hand, years of experience as a principal was not related to LC leadership. This may have been due to the fact that experienced principals were less motivated to engage in instructional leadership, especially if it had received little priority in their districts for much of their careers. For their part, less experienced principals may have been much more disposed to engage in LC leadership, particularly if they have been evaluated under systems that emphasized such leadership.

As noted, the findings produced by the HMLM model (Table 11) were consistent with those generated by the individual models. In particular, the measures for evaluation purpose and leadership activities assessed were both associated with a 0.15-unit increase in the extent to which principals engaged in LC leadership. In contrast, the measure for evaluation focus had a weaker impact on this outcome (i.e., a 0.05 unit increase). The study results suggest that district evaluation systems are likely to have a strong impact on LC leadership when they combine clear purposes related to learning and accountability (i.e., promoting principal PD and school restructuring, and holding leaders accountable) with an emphasis on leadership knowledge, skills, activities and impact on instruction and outcomes (i.e., goal setting, curriculum design, teacher PD and evaluation, and monitoring student performance).

Finally, the HMLM2 model (Table 12) examined whether district policy intentions (represented by district administrators' perceptions) were related to principals' views of how evaluation policy was actually enacted (i.e., their perceptions of evaluation practices). The finding that principals' perceptions of evaluation practices were not related to district intentions is consistent with earlier research (Davis & Hensley, 1999; Reeves, 2005; Thomas, Holdaway, & Ward, 2000). In this study, this finding may have been due to the fact that the district evaluation systems were too weak to significantly affect principals' perceptions (Goldring et al., 2008). Or, it may have been because district evaluation was disconnected with school leadership

practices. If districts are trying to link principal evaluation to promote learning-centered principalship, districts need to ensure that they are aware of evaluation purposes and how the evaluation foci and the leadership activities assessed in evaluation address these purposes.

CONCLUSION AND IMPLICATIONS

An extensive research base exists on effective school leadership (see reviews by Hallinger & Heck, 1996; Leithwood & Janzi, 2005). These studies envisioned several leadership behaviors that could lead to changes in school performance and improved student success. Drawing on the work of Murphy and his colleagues (2007), we refer to these behaviors in this study as aspects of learning-centered leadership. However, only a few studies have examined the impacts of district principal evaluation. For instance, prior studies on principal evaluation have employed qualitative methods to document evaluation systems, finding that evaluation practices typically vary across districts and that superintendents and principals usually view the purposes of evaluation differently (Davis & Hensley, 1999; Thomas, Holdaway, & Ward, 2000). Larger, survey-based studies have shown that evaluation criteria are often poorly specified and rarely attend to principals' ability to enact rigorous curricula or monitor instructional quality (Goldring et al., 2008; Reeves, 2005). And a few researchers have examined links between evaluation and aspects of principals' knowledge and leadership behaviors (Kimball, Milanowski, & McKinney, 2007; Smith, Munter, & Katterfield, 2008). However, few studies have investigated the characteristics of effective district principal evaluation systems and how evaluation can promote effective leadership activities.

The study described here builds on prior scholarship to examine the relationship between district evaluation practices and key outcomes for school leaders and to identify several characteristics of district evaluation that can promote learning-centered leadership. Specifically, we found that the purposes and foci of principal evaluation, and the leadership activities assessed, can have an important effect on the likelihood that school leaders will engage in learning-centered behavior. As districts consider ways to influence principals' behavior and promote student learning under No Child Left Behind and state accountability systems, these findings suggest ways they can reform or augment their systems for evaluating school leaders. First, the results indicate the need for districts to go beyond using evaluation results for personnel or salary decisions to employ them in determining principal professional development, promoting school restructuring efforts, and holding leaders accountable for student learning. Our findings demonstrated a relationship between districts establishing principal PD, restructuring, and accountability as the purposes of evaluation, and principals setting ambitious goals for student learning, coordinating curricula, and monitoring and supporting instructional quality.

Second, the results provide evidence that the activities addressed as part of evaluation matter greatly. When district evaluation systems are oriented toward instructional leadership activities, they have a strong effect on LC leadership in school goal setting, curriculum design, teacher PD and evaluation, and monitoring student learning. Finally, the focus of evaluation seems to affect the extent to which principals engaged in LC leadership. The findings suggest that concentrating (as part of evaluation) on principals' knowledge and skills, behaviors, professional relationships, and leaders' impact on school climate and student performance could make differences in LC leadership practices.

In terms of limitations, this study was unable to fully control for possible confounding variables because we used observational data and we were not able to randomly assign principals to different evaluation systems. While we included several variables to control for the effects of principal background and district contexts on the relationship between district evaluation practices and leaders' behaviors, we did not, for example, have measures of other district policies such as those related to student assessment, principal PD, or school reform. Further, as noted, the research took place in 13 school districts in Michigan, each of which had more than eight K–12 schools. As a result, the findings cannot be generalized to districts with eight or fewer schools or to a national sample. Therefore, one implication for future research would be to examine the same or similar questions by employing a randomized experimental design; such a design could include districts with various sizes in several states or a national sample. A second limitation, given the attention in the literature to new and more sophisticated approaches to principal evaluation, is that the 13 Michigan districts in this sample may not have been particularly innovative with regard to evaluating school leaders. Thus, in future studies, it could be useful to feature districts that have implemented systems based on the Danielson Framework (Danielson, 1996), the Interstate School Leaders' Licensure Consortium (ISLLC) standards (CCSSO, 1996), or the Vanderbilt Assessment of Leadership in Education (Murphy et al., 2007).

Third, the analysis presented here is limited by the fact that it did not include measures of teachers or students. Therefore, it would make sense for future research to include measures of teachers' instructional practices and student learning as part of HLM models. Such studies could provide insights into whether and how principal evaluation practices and principal leadership behaviors are related to instructional quality and student achievement. Finally, while this study documented relationships between principal evaluation purpose and focus, and the leadership activities assessed, and the extent to which school leaders engaged in LC leadership behaviors, we were not able to explain the processes by which these aspects of evaluation affected such behaviors. Therefore, it would be useful in future research to employ qualitative methods (including interviews, observations, and job

shadowing) to better understand how certain evaluation practices seem to contribute to desired leadership behaviors.

NOTES

1. The district's new principal evaluation system was based on its teacher evaluation system, which, in turn, was based on Danielson's Framework for Teaching (1996); the new system was characterized by clear performance standards (dimensions) and rubrics differentiating performance on the dimensions. Depending on a given principal's years of experience and prior evaluations, he or she was evaluated with regard to their performance on two or more of the dimensions each year. Under the new evaluation system, most of the data were collected by district administrators through observations and conferences. The district's old evaluation system, like those in many other districts, consisted of a checklist of personal traits and behaviors, and it did not feature a rubric differentiating levels of performance (Kimball, Milanowski, & McKinney, 2007).

2. In the study reported here, we did not include measures of professional community or relational trust in our models or analyses. This was due to the fact that our measures of these constructs were based on teacher surveys, which were not a focal part of this study. In other work, we investigate possible links between principal leadership behaviors and teachers' perceptions of professional community and relational trust in their schools.

3. The response rates were calculated by dividing the number of valid returned surveys by the total number of eligible participants (i.e., the total number of people who were initially contacted minus the number of respondents who were ineligible). With regard to the district administrator surveys, ineligible respondents were those who had not been responsible for principal evaluation in 2006–2007 in the same district (where they were working in 2007–2008). In terms of the principal surveys, ineligible respondents were those who had not worked as principal in 2006–2007 in the same school (where they were working in 2007–2008). With regard to the teacher surveys, ineligible respondents were those who had not worked as teachers in 2006–2007 in the same school (where they were working in 2007–2008).

4. The theory in this study supports the assumption that constructs (factors) are not independent. Therefore, we used oblique rotation to run the factor analysis, which relaxes the assumption of independent factors (constructs). The following aggregation used the same procedure.

5. Hierarchical Linear Modeling (HLM) is based on and makes use of nested data. Nested data can include, for example, repeated observations nested within persons (e.g., four dependent variables in this study can be considered as repeated observations of principal's leadership behaviors), persons nested within organizations (e.g., principals' work within school districts), and the organizational units themselves nested within larger entities (e.g., districts are located in states). The application of HLM addressed three general research purposes: (a) to improve estimation of effects within individual units (e.g., developing an improved estimate of a regression model for an individual principal by acknowledging that similar estimates exist for other principals within the district); (b) to improve the modeling of crosslevel effects (e.g., how principals' leadership influenced by district evaluation policy); and (c) to partition variance-covariance components (e.g., to decompose the variation among a set of individual-level variables into within- and between-school or district components). Please refer to Raudenbush and Bryk (2002) for further details.

REFERENCES

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement and Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Betts, J. R., & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22, 343–352.

- Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational Administration Quarterly*, 35(5), 751–781.
- Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.
- Burch, P., & Spillane, J. P. (2003). Elementary school leadership strategies and subject matter: Reforming mathematics and literacy instruction. *Elementary School Journal*, 103(5), 519–535.
- Carnoy, M., Elmore, R., & Siskin, L. S. (Eds.). (2003). *The new accountability: High schools and high-stakes testing*. New York: RoutledgeFalmer.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170.
- Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.
- Council of Chief State School Officers (CCSSO). (1996). *Interstate school leaders licensure consortium: Standards for school leaders*. Washington, DC: Author.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Davis, S. H., & Hensley, P. A. (1999). The politics of principal evaluation. *Journal of Personnel Evaluation in Education*, 13(4), 383–403.
- Desimone, L., Porter, A. C., Garet, M., Suk Yoon, K., & Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112.
- Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method*. Hoboken, NJ: John Wiley & Sons, Inc.
- Frank, K. A., Muller, C., Schiller, K., Riegle-Crumb, C., Strassman-Muller, A., Crosnoe, R., & Pearson J. (2008). The social dynamics of mathematics course taking in high school. *American Journal of Sociology*, 113(6), 1645–1696.
- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-income, low-achieving youth. *Educational Evaluation and Policy Analysis*, 19, 325–338.
- Garrett, W. R., & Flanigan, J. L. (1991). Principal evaluation: A definitive process. *Journal of School Leadership*, 1(1), 74–86.
- Goldring, E., Cravens, X. C., Murphy, J., Elliot, S. N., Carson, B., & Porter, A. C. (2008). The evaluation of principals: What and how do states and districts assess leadership? Paper presented at the annual meeting of the American Educational Research Association, New York.
- Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *The Elementary School Journal*, 96(5), 527–549.
- Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, 32(1), 5–44.
- Harrison, W. C., & Peterson, K. D. (1986). Pitfalls in the evaluation of principals. *Urban Review*, 18(4), 221–235.
- Hart, A. W. (1992). The social and organizational influence of principals: Evaluating principals in context. *Peabody Journal of Education*, 68(1), 37–57.

- Heck, R. (1992). Principals' instructional leadership and school performance: Implications for policy development. *Educational Evaluation and Policy Analysis*, 14(1), 21–34.
- Heck, R. H., & Marcoulides, G. A. (1992). Principal assessment: Conceptual problems, methodological problem, or both? *Peabody Journal of Education*, 68(1), 124–144.
- Kimball, S. M., Milanowski, A., & McKinney, S. A. (2007). Implementation of standards-based principal evaluation in one school district: First year results from randomized trial. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Kimball, S. M., & Pautsch, C. A. (2008). *Principal evaluation and support in two school districts using new leadership standards: A cross-site comparison*. Madison: University of Wisconsin-Madison.
- Knapp, M. S., Shields, P. M., & Turnbull, B. J. (1992). *Academic challenges for the children of poverty*. Summary report. Washington, DC: U.S. Department of Education.
- 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. (1994). Leadership for school restructuring. *Educational Administration Quarterly*, 30(4), 498–518.
- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research, 1996–2005. *Leadership and Policy in Schools*, 4, 177–199.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. New York: Wallace Foundation.
- Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education*, 106(4), 532–575.
- Louis, K. S., Marks, H. M., & Kruse, S. D. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33, 757–798.
- Mangin, M. M. (2007). Facilitating elementary principals' support for instructional teacher leadership. *Educational Administration Quarterly*, 43(3), 319–357.
- Marks, H. M., & Nance, J. P. (2007). Contexts of accountability under systemic reform: Implications for principal influence on instruction and supervision. *Educational Administration Quarterly*, 43(1), 3–37.
- Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39(3), 370–397.
- Murphy, J. F., Goldring, E. B., Cravens, X. C., Eliot, S. N., & Porter, A. C. (2007). The Vanderbilt assessment of leadership in education: Measuring learning-centered leadership. Nashville, TN: Vanderbilt University.
- Murphy, J., Hallinger, P., Peterson, K. P., & Lotto, L. S. (1987). The administrative control of principals in effective school districts. *Journal of Educational Administration*, 25(2), 161–192.
- Nettles, S. M., & Harrington, C. (2007). Revisiting the importance of the direct effects of school leadership on student achievement: The implications for school improvement policy. *Peabody Journal of Education*, 82(4), 724–736.
- Newmann, F. M. (1997). How secondary schools contribute to academic success. In K. Borman & B. Schneider (Eds.), *Yearbook of the National Society for the*

- Study of Education: Vol. 97, Pt. 1. The adolescent years: Social influences and educational challenges* (pp. 88–108). Chicago: National Society for the Study of Education.
- Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, 67(1), 41–74.
- Newmann, F. M., Smith, B. A., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23(4), 297–321.
- Peterson, K. D. (1984). Mechanisms of administrative control over managers in educational organizations. *Administrative Science Quarterly*, 29(4), 573–597.
- Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education*. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd edition). Thousand Oaks, CA: Sage Publications.
- Reeves, D. B. (2005). *Assessing educational leaders: Evaluating performance for improved individual and organizational results*. Thousand Oaks, CA: Corwin Press.
- Ross, J. A. & Gray, P. (2006). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education*, 29(3), 798–822.
- Smith, T. M., Munter, C., & Katterfeld, K. (2008). How do districts' leadership and teaching assessment tools/processes shape principals' visions of what counts as high quality instruction? Nashville, TN: Vanderbilt University.
- Snyder, J., & Ebmeier, H. (1992). Empirical linkages among principal behaviors and intermediate outcomes: Implications for principal evaluation. *Peabody Journal of Education*, 68(1), 75–107.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reforming and refocusing implementation. *Review of Educational Research*, 72(3), 387–431.
- Stein, M. K., & D'Amico, L. (2002). Inquiry at the crossroads of policy and learning: A study of a district-wide literacy initiative. *Teachers College Record*, 104(7), 1313–1344.
- Stein, M. K., & Spillane, J. P. (2005). How can educational leaders support and promote teaching and learning? New conceptions of learning and leading in schools. In W. A. Firestone & C. Riehl (Eds.), *A new agenda for research in educational leadership* (pp. 28–45). New York: Teachers College Press.
- Thomas, W. T., Holdaway, E. A., & Ward, K. L. (2000). Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education*, 14(3), 215–240.
- Weatherley, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Education Review*, 47(2), 177–197.
- Wong, K. K., Hedges, L. V., Borman, G. D., & D'Agostino, J. V. (1996). *Prospectus: Special analyses*. Final Report. Washington, DC: U.S. Department of Education.
- Youngs, P., & King, M. B. (2002). Principal leadership for professional development to build school capacity. *Educational Administration Quarterly*, 38(5), 643–670.

Copyright of Leadership & Policy in Schools is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.