American Educational Research Journal
http://aerj.aera.net

Assessing the Contribution of Distributed Leadership to School Improvement and Growth in Math Achievement

Ronald H. Heck and Philip Hallinger

Am Educ Res J 2009 46: 659
DOI: 10.3102/0002831209340042

American Educational Research Journal
September 2009, Vol. 46, No. 3, pp. 659–689
DOI: 10.3102/0002831209340042
© 2009 AERA. http://aerj.aera.net

Assessing the Contribution of Distributed Leadership to School Improvement and Growth in Math Achievement

Ronald H. Heck
University of Hawaii at Manoa
Philip Hallinger
Hong Kong Institute of Education

Although there has been sizable growth in the number of empirical studies of shared forms of leadership over the past decade, the bulk of this research has been descriptive. Relatively few published studies have investigated the impact of shared leadership on school improvement. This longitudinal study examines the effects of distributed leadership on school improvement and growth in student math achievement in 195 elementary schools in one state over a 4-year period. Using multilevel latent change analysis, the research found significant direct effects of distributed leadership on change in the schools' academic capacity and indirect effects on student growth rates in math. The study supports a perspective on distributed leadership that aims at building the academic capacity of schools as a means of improving student learning outcomes.

KEYWORDS: distributed leadership, collaborative leadership, school improvement, student learning, educational change

Over the past 40 years, researchers have sought to understand the contribution that leadership makes to effective schooling (Bossert, Dwyer, Rowan, & Lee, 1982; Firestone & Wilson, 1985; Gross & Herriot, 1965; Hallinger, Bickman, & Davis, 1996; Heck, Larsen, & Marcoulides, 1990; Marks

---

RONALD H. HECK is a professor of educational administration and policy in the College of Education/Educational Administration and Policy, University of Hawaii at Manoa, 1776 University Avenue, Honolulu, Hawaii 96822; e-mail: *rheck@hawaii.edu*. Research interests include organizational effectiveness, leadership, and policy effects.

PHILIP HALLINGER is chair and professor of leadership and change at the Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong; e-mail: *Hallinger@gmail.com*. His research interests are in leadership, school improvement, organizational change, and leadership development.

& Printy, 2003). Recent reviews of this literature suggest that substantial progress has been made in understanding both the extent of school leadership effects and the means by which leadership impacts school performance (Bell, Bolam, & Cubillo, 2003; Hallinger & Heck, 1996, 1998; Witziers, Bosker, & Kruger, 2003). One prominent observer recently concluded, "It has become increasingly clear that leadership at all levels of the system is the key lever for reform, especially leaders who a) focus on capacity building and b) develop other leaders who can carry on" (Fullan, 2006, p. 33).

These research findings have brought about two changes in the perspectives of educational researchers and policy makers. First, there is increased interest in how leadership is shared or "distributed" among administrators, teachers, and parents in schools (Gronn, 2002; Leithwood, Mascall, & Strauss, 2009; Spillane, 2006). Scholars now suggest that distributed leadership could provide a more sustainable means of building the type of learning-focused climate that characterizes high-performing schools (Day, Gronn, & Salas, 2006; Hallinger & Heck, 1998; Leithwood, Anderson, Mascall, & Strauss, in press; Leithwood, Louis, Anderson, & Wahlsttom, 2004; Spillane, 2006).

Second, there is focused interest in the role that leadership plays in bringing about school improvement over time (Leithwood et al., 2004; Luyten, Visscher, & Witziers, 2005; Reynolds, Teddlie, Hopkins, & Stringfield, 2000; Sleegers, Geijsel, & Van den Berg, 2002). Previous research has not adequately addressed the modeling of change in leadership, related educational processes, and student learning over time (Heck & Hallinger, 2005; Krüger, Witziers, & Sleegers, 2007). Modeling growth in achievement over time provides a more equitable means of assessing the contribution that schools make to the educational progress of students than simple comparisons of their outcome levels (Seltzer, Choi, & Thum, 2003).

In this article we ask, How does distributed leadership contribute to the improvement of learning in schools? We test a conceptual model in which the effects of distributed school leadership on growth in math achievement are mediated by the school's academic capacity and social-curricular organization. Our proposed analysis of leadership effects differs from previous quantitative work in this field through its focus on measuring organizational variables and student learning on multiple occasions and describing how changes in the initial levels of these organizational variables predict subsequent growth in student learning. Our focus on changes in these constructs over a 4-year period is intended to confirm a temporal sequence between school actions and student learning.

Our study extends earlier research on leadership and school improvement in two ways. First, despite calls for studies that examine policy prescriptions for shared leadership against empirical evidence, most studies have been descriptive rather than analytical (Heck & Hallinger, 2005; Leithwood et al., 2009; Pounder, Ogawa, & Adams, 1995). Our study tests a conceptualization of school leadership as an organizational property against empirical evidence of school improvement (Ogawa & Bossert, 1995).

Second, the growth modeling methods used in this study enable us to monitor changes among the constructs over time. Modeling growth trajectories provides a more accurate and thorough estimation of processes such as student learning than simple comparison of achievement levels at one point in time, learning gains between two measurements, or an achievement score adjusted for a previous score. Growth models incorporate more information about prior conditions than the other approaches (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). In growth models, both the level of outcomes attained and the rate of the change over time can be examined simultaneously. This may offer greater insight into how changes in distributed leadership contribute to growth in student learning.

Prior empirical research on school leadership effects consists almost exclusively of cross-sectional studies that describe these relationships at a single point in time (Hallinger & Heck, 1996; Krüger et al., 2007; Luyten et al., 2005; Southworth, 2002). This approach confounds the effects of time in relationships among variables (Davies, 1994) and is, therefore, ill-equipped to illuminate how leadership contributes to school improvement (Jackson, 2000; Leithwood et al., 2004; Ogawa & Bossert, 1995; Pounder et al., 1995). If we are to improve schools in a systematic way, then collecting high-quality information about school processes and outcomes over time is essential.

Leadership and School Improvement

One of the challenges of studying school leadership effects is the presence of multilevel organizational structures within educational organizations. Multilevel models of student learning assume that students are not randomly assigned to classrooms and that principals and teachers are not randomly distributed across schools (Lee & Bryk, 1989). Proposed models must account for how educational activities across multiple organizational levels subsequently influence the learning of individual students. There are obviously many indicators of schools' organizational and academic processes, as well as pathways of influence, and we caution that it would be a mistake to think of any subset used to explain student growth in learning as theoretically complete.

The phrase "school improvement leadership" implies the existence of a cause-effect relationship between the strategies of leaders, school improvement activities, teacher classroom practices, and growth in student outcomes. Although progress has been made in defining the nature of these relationships, scholars operating in the United Kingdom (Bell et al., 2003; Southworth, 2002, 2003), the United States (Bossert et al., 1982; Hallinger & Heck, 1996, 1998), Canada (Leithwood et al., 2004, in press; York-Barr & Duke, 2004), the Netherlands (Krüger et al., 2007; Sleegers et al., 2002; Witziers et al., 2003), and AnZed (Mulford & Silins, 2003; Robinson, Lloyd, & Rowe, 2008) continue to debate the meaning of empirical findings on school leadership effects.[1] Moreover, the predominant assumption that leadership impacts school improvement understates the extent to which leaders are influenced by the organizational environment (Hallinger & Heck, 1996; Krüger et al., 2007; Leithwood et al., 2004; Southworth, 2002). Thus, we conclude that research on school leadership effects must take into account features of the organizational context and continue to approach issues of causal inference with caution.

Sources of Leadership

The study of school leadership must be explicit about the sources of leadership. Although prior research has generally highlighted the leadership role of the principal, this study focuses on distributed leadership (Gronn, 2002; Harris, 2003; Spillane, 2006). This refers to forms of collaboration practiced by the principal, teachers, and members of the school's improvement team in leading the school's development.

The rationale for distributed school leadership is grounded in the concept of sustainable change (Fullan, 2001). Leadership must create changes that are embraced and owned by the teachers who are responsible for implementation in classrooms (Fullan, 2006; Hall & Hord, 2001). Moreover, given the intensification of work activities of school administrators, selected approaches to leadership must also be sustainable for those who lead (Barth, 1990; Donaldson, 2001). As Hall and Hord conclude, "Principals can't do it alone." Thus, scholars assert that sustainable school improvement must be supported by leadership that is shared among stakeholders (Barth, 2001; Fullan, 2001; Harris, 2003; Marks & Printy, 2003; Stoll & Fink, 1996).

Means of Leadership

We define school improvement leadership as an influence process through which leaders identify a direction for the school, motivate staff, and coordinate an evolving set of strategies toward improvements in teaching and learning. This emphasizes our belief that the effects of school leadership are largely mediated by academic and social conditions present in the school and aimed toward learning outcomes. Empirical evidence, though not conclusive, does provide insight into the means by which leadership impacts teaching and learning. Specifically, we find that school improvement leadership:

- Impacts conditions that create positive learning environments for students (Geijsel, Sleegers, Stoel, & Krüger, 2009; Hallinger et al., 1996; Hallinger & Heck, 1998; Heck et al., 1990; Leithwood et al., 2004, in press; Robinson et al., 2008; Sleegers et al., 2002; Wiley, 2001).
- Mediates academic expectations embedded in curriculum standards, structures, and processes as well as the academic support that students receive (Cohen & Hill, 2000; Darling Hammond, 2006; Hallinger et al., 1996; Hill & Rowe, 1996; Lee & Bryk, 1989; Oakes, 2005).
- Employs improvement strategies that are matched to the changing state of the school over time (Jackson, 2000; Leithwood et al., 2004, in press; Mulford & Silins, 2003; Reynolds et al., 2000; Stoll & Fink, 1996).

- Supports ongoing professional learning of staff, which, in turn, facilitates efforts of schools to undertake, implement, and sustain change (Barth, 1990; Crandall, Eiseman, & Louis, 1986; Fullan, 2006; Geijsel et al., 2009; Hall & Hord, 2001; Robinson et al., 2008; Stoll & Fink, 1996).

This description of the means by which leadership impacts school improvement is consistent with what scholars have termed a *mediated-effects* model of leadership (Baron & Kenny, 1986; Pitner, 1988). Leadership effects on learning are brought about indirectly through their impact on people, structures, and processes in the school over time (Bossert et al., 1982; Hallinger & Heck, 1996; Leithwood et al., in press).

Modeling Distributed Leadership Effects on Student Learning

Structural equation modeling (SEM) is often used to test the plausibility of proposed theoretical relationships between variables in nonexperimental research (Jöreskog & Sörbom, 1989). The procedure can be formulated as an estimation of the coefficients of a set of simultaneous equations representing the proposed relationships. The analysis involves imposing a set of model restrictions on the sample covariance matrix and trying to determine whether the proposed set of model restrictions, or an alternative set, fits best in the population under study. In this study, we employ multilevel latent change analysis (LCA), an application of SEM for modeling longitudinal data, to examine changes in leadership, school academic capacity, and student math outcomes over a 4-year period. In LCA, developmental processes can be conceptualized as continuous latent intercept (e.g., initial status) and growth factors (Muthén & Muthén, 1998–2006). Repeated measures data (e.g., math tests, survey items) serve as the observed indicators of the underlying factors. In SEM path diagrams, such as Figure 1, the latent factors are delineated by ovals, and their sets of observed indicators are delineated by rectangles. The underlying factors are assumed to explain the pattern of covariation in the sample covariance matrix observed between the indicators. This is represented in the figure by arrows from the latent factors to their observed indicators.

After measuring the latent change processes using the sets of repeated observed measures, the second part of LCA involves examining the structural relationships between the latent change factors and other observed and latent variables in the model. In formulating structural models, researchers often draw a distinction between *exogenous* variables (i.e., variables whose variability is accounted for by factors outside the model) and *endogenous* variables (i.e., variables whose behavior is dependent upon other variables within the model). The goal is to solve the equations for the endogenous variables taking into account the exogenous variables and random errors between constructs. In Figure 1, we represent the exogenous variables as unshaded rectangles or ovals. We represent the endogenous factors as shaded ovals and their observed indicators as shaded rectangles.

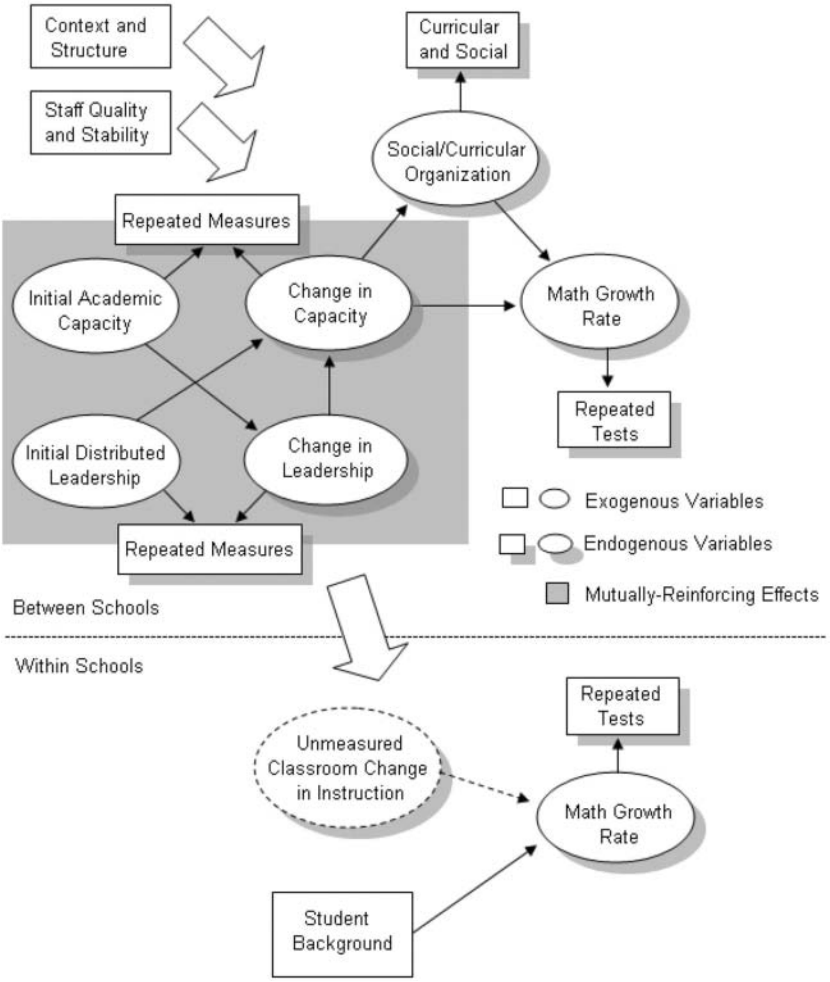

Figure 1. **Conceptual Model of School Improvement Leadership and Student Learning**

Exogenous Variables

Within schools, student background represents a set of exogenous variables that is proposed to explain a portion of their growth in math. At the school level, the model includes several exogenous variables that previous research has identified as affecting student achievement. These variables include school size, student composition, as well as teacher professional preparation, certification, and stability (Goldhaber, 2002; Leithwood et al., 2004; Southworth, 2002; Teddlie, Stringfield, & Reynolds, 2000). In addition to its direct effect on achievement, student composition (e.g., social class, race/ethnicity, language background) has been found to affect academic expectations, curriculum organization, grouping, teacher behavior (Lee & Bryk, 1989; Oakes, 2005), and school leadership (Hallinger & Murphy, 1986). Features of small schools appear to favor enhanced growth in student learning (Leithwood et al., 2004; Mulford & Silins, 2003; Southworth, 2003). In Figure 1, the large arrow from context and structure indicates that we expect these to influence other variables in the model, although we do not hypothesize their specific effects in the study.

Staffing variables such as teacher certification and stability are potentially important variables because previous research has found that schools in communities serving concentrations of low SES and students of color can experience greater difficulty hiring and retaining quality faculty and administrators (Darling Hammond, 2006; Goldhaber, 2002). School inequities (e.g., resources, personnel, turnover) can compromise the quality of student learning outcomes (Oakes, 2005). These types of indicators may moderate schools' strategic efforts to improve educationally. Previous research also suggests that principal stability can influence the management of school improvement projects (e.g., Firestone & Wilson, 1985). In Figure 1, the large arrow from staff quality and stability suggests that staffing variables will affect the endogenous constructs in the model positively (e.g., sociocurricular organization, change in achievement, change in academic capacity).

In Figure 1, we also note that the initial status leadership and academic capacity latent factors are represented as exogenous variables (shown as unshaded ovals). This is because within the 4-year temporal sequence implied by our proposed model, their variability is assumed to be determined by prior relationships outside the model.

Endogenous Variables

In the proposed model, the endogenous variables serve as mediating organizational processes between the exogenous variables and growth in student math outcomes. We conceptualize four endogenous latent constructs (measured by sets of observed indicators): change in distributed leadership, change in school academic capacity, sociocurricular organization (measured in Year 4 of the study), and change in math achievement. The first, change in distributed leadership, has been discussed in the previous section.

The second variable, change in school academic capacity, refers to changes in conditions of the school that support the provision of effective teaching and learning and enable the professional learning of the staff (Darling Hammond, 2006; Hallinger & Heck, 1998; Hill & Rowe, 1996; Robinson et al., 2008; Stoll & Fink, 1996). In Figure 1 we propose that change in distributed leadership will positively affect change in academic capacity directly and sociocurricular organization and growth in student learning indirectly. We also hypothesize that change in academic capacity will positively affect school sociocurricular organization and growth in student learning. We highlight this leadership-academic capacity portion of the model in gray in order to emphasize our focus on these particular constructs as representing a mutually reinforcing process.

Variables are mutually reinforcing if each leads to change in the other (Marsh & Craven, 2006). More specifically, the leadership-academic capacity portion of the model represents two parallel growth processes (see Muthén & Muthén, 1998–2006). Our model therefore implies that the leadership and capacity-building growth trajectories found in schools (and the math achievement trajectories of individual students within schools) have common algebraic forms but that not every school has the same trajectory (Singer & Willett, 2003).

A third mediating factor is the school's sociocurricular organization (Lee & Burkam, 2003; McCaffrey et al., 2004). An extensive literature describes how schools' sociocurricular organization impacts student learning opportunities and educational attainment (e.g., Alexander & Cook, 1982; Braddock & Slavin, 1993; Burns & Mason, 1998; Lee & Bryk, 1989; Oakes, 2005). In models of school effects, the sociocurricular organization of the school mediates between contextual (e.g., social composition) and structural conditions (e.g., enrollment, type of school) and student outcomes (Lee & Burkam, 2003). Lee and Burkam (2003) define curricular organization as students' access to quality curricular experiences within the school. Social organization refers to the pattern of social relationships among administrators, teachers, and students (e.g., presence of supportive relationships, student integration, and well-being). Within classrooms, individual students benefit from positive relationships with teachers (Fullan, 2001; McCaffrey et al., 2004). At the school level, patterns of teacher-student interactions tap into the quality of sociocurricular organization (Lee & Burkam, 2003; Oakes, 2005).

If academic capacity is a key target of leadership efforts designed to impact teacher practice and student performance, then, as Figure 1 suggests, we propose that changes in school academic capacity should be reflected in student perceptions of the school's classroom curriculum and social relationships between students and teachers. Moreover, we propose students' perceptions of their sociocurricular relationships with teachers will be positively related to growth in achievement.

The fourth endogenous variable is math achievement (measured by repeated state tests). We represent student growth in math at two organizational levels (i.e., the student and school levels) and propose that students' growth in math is a parameter that varies randomly across schools. This implies that student growth rates are different within the population of schools. The subsequent objective is to explain this variability in growth rates through the contextual and organizational variables proposed in the model.

Classrooms represent an organizational level that mediates the effects of schoolwide improvement activities on individual student progress. We note that this multilevel study does not include a direct measure of change in the instructional practices of teachers (shown as a dotted, shaded oval in Figure 1). We acknowledge that classroom-level information would be desirable in order to provide a more complete picture of the organizational processes at work in these schools; however, such data are exceedingly difficult to obtain. Indeed, none of the often-cited school leadership studies has included such data (e.g., Hallinger et al., 1996; Heck et al., 1990; Leithwood & Jantzi, 1999; Marks & Printy, 2003; Pounder et al., 1995; Wiley, 2001).

Nonetheless, we must be explicit that we assume that changes in school leadership and capacity building processes exert "trickle down" cross-level effects on teacher classroom behavior. This is indicated in Figure 1 by a large arrow from the latent change factors to classroom changes in instruction. We represent this construct as a dotted latent variable in the figure to emphasize we do not have direct observed measures of classroom changes. Over time, however, classroom instructional differences contribute to variability in student growth rates (Cohen & Hill, 2000; Creemers, 1994; Heck, 2009), represented by a dotted arrow from changes in classroom practices to student growth rates. Although the inability to test this assumption directly represents a limitation of the study, we do report the results of comparing teacher perceptions of changes in classroom practices against students' perceptions of the same classroom changes. In related research conducted on teacher data, we also determined that differences in the effectiveness of successive classroom teachers (accounting for about 11% of the total variability in math outcomes) and the school's collective teaching effectiveness contribute meaningfully to reducing gaps in student learning in math between schools. We place this design limitation in perspective in the concluding section of the article.

Research Questions

We propose two broad research questions in this study. The questions are framed within the conceptual model proposed above and portrayed in Figure 1.

What is the relationship between distributed leadership and academic capacity when observed over time? We assert that school improvement represents a dynamic process that involves changes in the state of the organization over time. Our model proposes that changes in distributed leadership and academic capacity represent a mutually reinforcing process. Initial distributed leadership is proposed to be positively related to change in school academic capacity and initial academic capacity to change in distributed leadership.

How does distributed leadership impact school improvement capacity and subsequent growth in math? The second question seeks to illuminate how changes in levels of distributed leadership and academic capacity carry over to changes in math achievement. We propose that school academic capacity and sociocurricular organization function as mediators between distributed leadership and student growth. We assess the strength of the mediated effects (and indirect leadership effects) in accounting for growth in student learning (Calsyn, Winter, & Burger, 2005). We test several propositions related to this question.

First, we propose that change in distributed leadership will be *directly* and significantly related to change in academic capacity. Second, we propose that changes in academic capacity will be *directly* and significantly related to (a) growth in student learning and to (b) student perceptions of sociocurricular organization. Third, we propose that change in distributed leadership will be *indirectly* and significantly related to change in sociocurricular organization and math achievement. Finally, we propose that change in distributed school leadership will be contingent on student composition and principal stability.

Research Method, Data, and Measures

This study employs a longitudinal nonexperimental design (Campbell & Stanley, 1966). Longitudinal nonexperimental studies are often used to study developmental trends (Marsh & Craven, 2006). Although superior to cross-sectional designs when temporal relationships are a focal point of the analyses, they do not fully resolve issues of causal direction between variables (Cook, 2002). The major threat to validity in longitudinal nonexperimental research lies in uncontrolled or confounding variables.

To test the model, data were collected from students and teachers in elementary schools in a western state in the United States over a 4-year period. We captured changes in school processes through surveys given to each school's teachers on three occasions (Years 1, 3, and 4). Return rates for the three periods were 73.4% (n = 3,911), 78.6% (n = 4,152), and 76.2% (n = 4,055), respectively. The survey is administered at regular cycles in each school to all certified staff, Grade 5 students, and a random sample of parents (i.e., about 20% across grade levels in each school). Where surveys are repeated over time with a high level of consistency between items, the measures may be used to estimate changes in a population (i.e., referred to as a longitudinal panel study; Davies, 1994). Achievement data from a student cohort were collected in Years 2, 3, and 4 (i.e., corresponding to their third-, fourth-, and fifth-grade years). Unequal spacing of observations and nonlinearity can be incorporated into an LCA model without compromising quality of data analysis (Raykov & Marcoulides, 2006).

Data

Data were from a random sample of public elementary schools (n = 195). From these schools, participating students were drawn from a third-grade student cohort (n = 13,389) that was subsequently observed over a 3-year period. Background data were as follows: female, 49%; participation in federal free or reduced lunch program, 45%; receiving English language services, 7%; receiving special education services, 11%; minority, 50%; and changed schools, 16%. One of the advantages of growth modeling is that missing data (i.e., less than 5%) and student mobility can be incorporated directly into the analysis, which reduces parameter bias that would result from eliminating these students (Peugh & Enders, 2004).

Measures

The theoretical model described earlier was operationalized through explicit measurement of the exogenous and endogenous variables included in Figure 1.

Background and context variables. Background variables included female (coded 1, male coded 0), low socioeconomic status (i.e., participation in the federal free or reduced lunch program coded 1, otherwise coded 0), special education services (coded 1, otherwise coded 0), minority by race/ethnicity (coded 1, otherwise coded 0), English language learning (ELL) services (coded 1, otherwise coded 0), and changed schools (coded 1, otherwise coded 0).

At the school level, context and structural indicators describe schools during the first year of the study (2002–2003). Means and standard deviations are provided in parentheses (not tabled). *Student composition* was defined as a composite variable by combining several relevant demographic indicators to create a weighted school indicator ($M = 0.00$, $SD = 1.00$). Larger positive values represent schools where percentages of these students were higher. The variables included percentage of children receiving free or reduced lunch ($M = 50.5$, $SD = 22.6$), percentage of students receiving ELL services ($M = 8.5$, $SD = 9.2$), and the percentage of racial/ethnic minority students ($M = 51.2$, $SD = 24.0$). *School size* was defined as the number of students enrolled for the school year ($M = 495.88$, $SD = 243.67$).

Staffing variables. Teacher quality was defined as the percentage of teachers at each school who met No Child Left Behind (NCLB) and state teacher licensing criteria ($M = 84.1$, $SD = 7.2$). *Teaching staff stability* was defined as the percentage of teachers in each school who had been at the school for 5 years ($M = 60.21$, $SD = 14.1$). We measured both of these variables during Year 4 of the study. Since NCLB was implemented, the state has tracked percentages of fully qualified teachers. Data on teacher qualifications between 2003 and 2006 suggest that local teacher labor market conditions continued to necessitate hiring considerable percentages of teachers who were less than fully qualified.[2] *Principal stability* was defined as whether the same principal (coded 1, otherwise 0) was at the school during the 4 years of the study ($M = 31\%$).

Distributed leadership and academic capacity. For the purposes of this study, information from three successive teacher surveys was used to measure these two constructs. Observed indicators were measured by 5-point, Likert-type scales. Higher means reflect stronger agreement with the items defining each subscale. Cronbach's (1951) alpha, a measure of internal consistency, was used to assess the reliability of each subscale.

Distributed leadership was measured by a composite set of items describing teacher perceptions of leadership exercised from a variety of sources within the school ($\alpha = .82$). The stem used for these items was, "To what extent does school leadership . . . ." The state survey items were designed to reflect three specific aspects of distributed leadership within each school (with items paraphrased in parentheses):

Make collaborative decisions focusing on educational improvement (i.e., ensure teachers have a major role in decisions about curriculum development in the school; enable administrators, teachers, and staff work together effectively to achieve school goals);

Emphasize school governance that empowers staff and students, encourage commitment, broad participation, and shared accountability for student learning (i.e., provide opportunities for parents to participate in important decisions about their children's education through a variety of venues; ensure teachers can freely express input and concerns to the administrators; provide opportunities for teachers to plan and make school decisions); and

Emphasize participation in efforts to evaluate the school's academic development (i.e., ensure adequate resources are available to the school to develop its educational programs; provide regular opportunities for all stakeholders to review the school's vision and purpose).

Observed leadership scores for each measurement occasion were used to define the leadership factor (see Figure 2). Positive growth in leadership over time results when teachers assigned higher scores on the leadership subscales at successive intervals.

School academic capacity ($\alpha = .94$) was measured by four subscales:

Standards emphasis and implementation ($\alpha = .91$): School's educational programs are aligned to the State content and performance standards; Teaching and learning activities are focused on helping students meet the State content and performance standards; School prepares students well for the next school; Students and parents are informed about what students are expected to learn; School has high academic and performance standards for students; Classroom instruction includes active participation of students; Curriculum and instructional strategies emphasize higher level thinking and problem solving; Instructional time is flexible and organized to support learning; Teachers provide a variety of ways for students to show what they have learned; Students learn to assess their own progress and set their own learning goals; Students are provided with multiple ways to show how well they have learned; Homework assignments are appropriate, productive, and reflective of adopted learning standards; Assessment results are used to plan and adjust instruction.[3]

Focused and sustained action on improvement ($\alpha = .83$): The school clearly communicates goals to staff, parents, and students; Vision and purpose are translated into appropriate educational programs for children; School seeks ways to improve its programs and activities that promote student achievement; Teachers know what the school learner outcomes are; Teachers expect high-quality work; School's vision is regularly reviewed with involvement of all stakeholder groups; Changes in curriculum materials and instructional practices are coordinated schoolwide; I am involved in the school improvement process.

Quality of student support (α = *.85):* Standards exist for student behavior; Discipline problems are handled quickly and fairly; School environment supports learning; Open communication exists among administrators, teachers, staff, and parents; Teachers feel safe at school; Teachers and staff care about students; Administrators, teachers, and staff treat each other with respect; I provide students with extra help when they need it; Programs meet special needs of students; School reviews support services are offered to students.

Professional capacity of the school (α = *.80):* Teachers are well qualified for assignments and responsibilities; Leadership and staff are committed to school's purpose; Staff development is systematic, coordinated, and focused on standards-based education; Systematic evaluation is in place.

Preliminary analyses determined how well the four indicators defined the latent academic capacity factor at each measurement occasion. Factor loadings across occasions averaged .94, .91, .96, and .92, respectively (not tabled). This suggests the scales were strong measures of the underlying academic capacity factor. Factor scores for each occasion were saved as variables and used in subsequent analyses. Positive growth in capacity over time means that teachers assigned higher scores on the subscales comprising academic capacity at succeeding occasions.

Sociocurricular organization. Sociocurricular organization was defined by fifth-grade student perceptions of the quality of their social relationships with teachers (and other adults) in the school and their experience of academic-curricular processes. We obtained these measures in Year 4 (return rate = 91%). The items comprising the two subscales were as follows:

Social organization (α = *.92):* I can freely express my opinions or concerns to my teachers; I can talk to my teachers, counselors, or other adults at school when I need to; My teachers care about me and treat me with respect; Students get along with each other pretty well at my school; My teachers give me extra help when I need it; I get help from the counselor when I need it; I enjoy coming to school.

Curricular organization (α = *.94):* School work is challenging; What I am learning will help me in the next grade; The programs at my school are good; What I am learning helps me reach the content and performance standards; My homework assignments help me to learn better; My teachers teach me how to think and solve problems; Most of my teachers teach in a way that is clear and easy to understand; My teachers make learning interesting in different ways; If I am having trouble learning something, my teachers usually find another way to help me understand it; We learn by doing things, not just by sitting and listening; I have learned to evaluate my own work and keep track of my progress; Students can show what they have learned in different ways—projects,

> portfolios, presentations; My teachers tell me how I am doing and how I can improve; I am aware of how well I am doing in class; My teachers discuss my progress in class with me on a regular basis; My teachers explain to me what they want me to learn; My teachers expect me to do quality work.

Math achievement. The math test used in the study was constructed to measure State-developed math content standards. The test consisted of constructed-response items and standardized test items from the Stanford Achievement Test (Edition 9). The test assesses student learning in five strands (number and operation; measurement; geometry and spatial sense; patterns, functions, and algebra; and data analysis, statistics, and probability) consisting of 52 items. Student scores (rescaled to range from 100 to 500) considered patterns of right, wrong, and omitted responses over successive years and were equated across the 3 years to enable the measurement of academic growth.

Data Analysis

Our proposed model highlights several features of data that must be incorporated into the analysis. First, the analysis must reflect the multilevel, nested structure of schools (Ogawa & Bossert, 1995). Accurate estimation of school parameters requires adjustment for the clustering of students within schools (Hill & Rowe, 1996). Second, repeated observations describing changes in individual students or changes in schools over time also represent nested data structures. This requires an analytic approach capable of incorporating changes in several variables at multiple organizational levels in one simultaneously estimated model (Singer & Willett, 2003). Third, longitudinal models require the specification of a temporal sequence of relationships among variables. In our study, relationships between prior and subsequent conditions are conceived as dynamic and possibly mutually reinforcing (Marsh & Craven, 2006). Our approach to multilevel, longitudinal modeling enables representation of initial states of variables and subsequent changes that occur between them over time. Fourth, in the context of testing proposed structural equation models, we recognize the need to consider alternative explanations and interpretations of our findings. Caution should therefore be exercised in using SEM applications to test substantive theories. Omitted variables and measurement error are common sources of model misspecification that can produce misleading results (Bentler & Bonett, 1980). More specifically, there may be unmeasured exogenous or endogenous variables that may be correlated with major constructs, such as leadership, in our model. These could compromise the validity of our proposed model. We next describe some of the steps we took to lessen this likelihood.

Preliminary analyses. We conducted several preliminary analyses to investigate possible relationships between exogenous school and staffing indicators that might influence our results (not tabled). For example, we found teaching staff stability was positively, but weakly, correlated with

Table 1
Model Fit Indices

Model	χ^2	*df*	CFI	RMSEA	R^2
1. Variance components model for latent constructs	91.55	32	.994	.012	0.00
2. Context and staff indicators	60.48	18	.995	.013	0.75
3. Context, staff, leadership, and school capacity	167.12	38	.989	.016	0.86
4. Complete Figure 1 model	380.39	95	.980	.015	0.88
Tests of single paths in Figure 1 model					
5. Path from change in academic capacity to change in leadership	392.01	95	.980	.015	0.88
6. Figure 1 model, with direct leadership effect	380.78	94	.980	.015	0.88

Note. CFA = Comparative Fit Index; *df* = degrees of freedom; RMSEA = Root Mean Square Error of Approximation Index.

principal stability ($r = .18$, $p < .05$) and the percentage of fully qualified teachers comprising the staff ($r = .32$, $p < .05$). Because principals can exercise influence in hiring teachers, we also investigated how school conditions might influence patterns of teacher mobility over time. We estimated that teacher turnover averaged about 8% per year during the years of our study. We also noted that the total set of school (i.e., student composition, student achievement, enrollment size) and staffing conditions (i.e., teacher, principal stability, teacher experience) included in our model contributed little (about 1%) in explaining school variability in the mobility of teachers over the years of the study (not tabled).[4]

Testing the proposed model. We next turned our attention to testing our proposed model in several steps. The indices describing the fit of each model to the data are summarized in Table 1. In LCA, repeated observations on individuals over time (y_t) can be expressed as a measurement model where the intercept and growth latent factors are measured by the multiple indicators of y (see note for further details).[5] The intercept factors representing the constructs were defined to represent initial achievement, leadership, or academic capacity, which is accomplished by setting each factor loading to 1.0 (as shown in Figure 2). The growth factors were defined to incorporate possible nonlinearity in the growth trajectories. This was accomplished by fixing the factor loading for the first measurement occasion to 0 and the second occasion to 1 and by letting the coefficient for the third occasion be estimated by the software. The estimate for this latter coefficient is indicated by an asterisk in Figure 2, which also indicates its statistical significance. The size and statistical significance of the estimated factor loading determine the shape of the growth trajectory (Raykov & Marcoulides, 2006).

Partitioning the variance in growth into its within- and between-group components is an important first step in determining whether a multilevel analysis is justified. This first model (Model 1 in Table 1) does not include predictors. If sufficient variance in growth rates exists between schools (e.g., over 5%), a school-level model can be developed to explain variability in this portion of the outcome. Our "variance components" model also summarizes the means and variability in the other endogenous factors.

Model 2 investigated the direct relationships between the context variables and math growth. Student-level variables were centered on their grand means. This results in school means that are adjusted for differences between students. This adjustment provides a more equitable comparison between schools in terms of what they contribute to growth in student learning. School-level estimates were also centered on their grand means (except for the dichotomous indicator of principal stability).

Model 3 added the mediating distributed leadership and academic capacity growth factors to the model. Model 4 added the mediating sociocurricular organization latent variable to the model. Adding this mediating variable to the model allowed us to examine whether the school organization construct might diminish or eliminate any direct effect of change in academic capacity (and indirect effect of leadership) on student growth in math. For example, if this variable eliminated the influence of the other key change constructs on school growth rates, it would invalidate our proposed model of school improvement. The between-school estimates for Model 4 are summarized in Figure 2.

Investigating specific propositions. Finally, we tested the validity of our model by examining two specific propositions about paths in the model. These tests are also summarized in Table 1. More specifically, we investigated an alternative model (Model 5) with a structural path from change in academic capacity to change in leadership instead of from change in leadership to change in capacity (as in Model 4). We also compared the fit of our proposed model of indirect leadership effects on math outcomes (through academic capacity) against a more general model that proposed both a direct effect and an indirect leadership effect on outcomes (Model 6). Such tests explore the validity of a proposed model, and in this case, the test was conducted by estimating an additional path between change in leadership and growth in math. We then compared the subsequent change in chi-square ($\Delta\chi^2$) between the two models after appropriate scaling adjustment for nonnormality (Muthén & Muthén, 1998–2006).

Results

Evaluating Alternative Models

Tests of our proposed model were conducted with Mplus 5.2 (Muthén & Muthén, 1998–2006). We present results based only on the series of models we originally proposed. In testing models using SEM, the emphasis is on

Table 2

Descriptive Statistics for Between-School Latent Variables in Model 1

	M	*SD*	Variance	Sig.*
Initial math level	217.10	38.59	1488.820	.000
Average math growth	16.58	17.61	310.112	.000
Initial leadership	–0.01	0.14	0.020	.000
Growth in leadership	0.00	0.05	0.003	.049
Initial academic capacity	–0.03	0.99	0.989	.000
Growth in capacity	0.01	0.59	0.342	.001
Social/curricular organization	–0.05	0.22	0.050	.000

Note. Sig.* = significance level for a test of whether the variance differs across schools.

specifying a set of theoretical relationships before testing them against the data. The goal is to reproduce the original matrix of covariance relationships with a set of model-proposed restrictions placed on it. In testing models, if a proposed model does not fit the data well, it would have to be reconceptualized. In contrast, if a proposed model fits the data well, this implies that it is a plausible representation of the data, but it may not be the only plausible representation (Hoyle & Panter, 1995).

In practice, one may not have only one model (or set of restrictions) in mind but rather a series of competing models. Testing the adequacy of each proposed model in sequence is known as an alternative-models approach (Hoyle & Panter, 1995). Through these comparisons, one can determine whether the alternative models fit the data as well, better, or worse than the primary model.

Models are evaluated in terms of their substantive features and the adequacy of their fit against the data. The adequacy of fit of each proposed model to the data, summarized in Table 1, was determined by several model fit indices. Although the chi-square statistic is often used in evaluating models, it has the undesirable property of being affected by sample size. With large samples, this can lead to falsely rejecting proposed models that otherwise fit the data quite well (Raykov & Marcoulides, 2006). In order to address this limitation, we also report the Root Mean Square Error of Approximation (RMSEA) fit index and the Comparative Fit Index (CFI; Raykov & Marcoulides, 2006). RMSEA describes the amount of model discrepancy per degree of freedom in the model. Values near .05 or lower generally indicate an adequate fit of the model to the data. The CFI compares the fit of the proposed model against a baseline (nonfitting) model, with values of at least .95 providing evidence of an adequate model fit.

The model fit criteria in Table 1 suggest each proposed model provided an adequate fit to the data (i.e., CFI values above .95; RMSEA values below .02). The table also provides an estimate of the variance in student growth accounted for by each model that includes predictors (Models 2–4). Model 2, which consisted of context and staffing variables and initial achievement status, accounted for about 75% of the between-school variance in math growth. Model 3, which added the mediating leadership and academic capacity factors, accounted for an additional 11% of the variance. Model 4, which added the social and curriculum organization construct, accounted for an extra 2% of variance in growth (total R^2 = 88%). Thus, Models 2 to 4 accounted for substantial amounts of variance in school math growth.

Because model testing revealed a strong fit to the data, we can turn our attention to the specific parameter estimates in our proposed models. We begin by discussing the means and variability in the endogenous factors as specified in our variance components model (Model 1). In Table 2, between schools, the math growth variance component was 310.11. Within schools, the factor variance component was 2,110 (not tabled). This suggests about 13% of the variability in latent growth lies between schools [310/(2,110 + 310)]. This result implies that the proposed model may be useful in explaining differences in math growth rates between schools. Data in the table further suggest there was significant variance in math growth ($\sigma^2_M = 310.11$, $p < .01$) between schools.

Turning to the leadership and academic capacity factors, the results in Table 2 indicate the initial (Year 1) leadership factor mean was −.01, and the variability in means was significant ($\sigma^2_{IL} = .020$, $p < .01$). The leadership growth slope was .00 (i.e., −.004), which because of our coding scheme, can be interpreted as the average change in leadership between the first and second intervals of the study (i.e., Year 1 and Year 3). This result suggests there was little average change in leadership over this period. However, there was significant variability in leadership growth slopes across schools ($\sigma^2_{SL} = .003$, $p < .05$). The initial academic capacity factor mean was −.03, and the mean academic capacity growth slope was .01. Again, this result suggests there was little average change in the academic capacity of schools between the first and second intervals of data collection. At the same time, however, both average initial academic capacity levels and average change in capacity varied significantly across schools ($p < .01$). These results indicate there was observed variation in the leadership and academic capacity change trajectories of individual schools.

Distributed Leadership Effects on School Improvement

The next portion of the analysis focuses on explaining variability in the endogenous factors. Model 4 (in Table 1) summarizes our proposed Figure 1 model with indirect effects of distributed leadership on growth in math. Because Model 4 fits the data well, in Figure 2, we summarize the effects of the school-level exogenous and endogenous variables on math growth. We first examine the trajectories for the latent growth factors. For math growth, the third occasion in Figure 2 was estimated as 2.0 ($p < .05$). This suggests a linear growth trajectory, since the growth rate between each occasion was 1.0. In contrast, for growth in capacity, the third occasion was estimated as

1.6 ($p < .05$), which suggests decelerating growth between the second and third occasions, compared with the first and second occasions. For change in leadership, the third estimate was 4.1 ($p < .05$), which suggests the trajectory had a quadratic shape (i.e., increasing between Occasions 1 and 2 but declining between Occasions 2 and 3).

We note in passing that all of the student background variables were significantly related to *growth rates* in math.[6] The coefficients in Figure 2 are standardized, which indicates the relative size of each variable's effect (the significance level was set at $p = .05$). When interpreting effect sizes, the level of analysis matters in multilevel populations. For example, a standardized effect that is small in accounting for existing variation at the student level (e.g., 0.1 or 0.2) may be large in accounting for between-school variation (Bloom, Hill, Black, & Lipsey, 2008). Between schools, students' yearly math growth rate was about 0.5 of a standard deviation (not tabled). A between-group effect of 0.2, therefore, would increase the yearly growth rate by about 40% (Bloom et al., 2008). It is therefore best to consider specific effects in relation to others at each level of the model.

With respect to the school context, structure and staffing variables, student composition (standardized $\gamma = -.20$, $p < .05$), teacher professional preparation (standardized $\gamma = .12$, $p < .05$), and school size (standardized $\gamma = -.10$, $p < .05$) were significantly related to math growth rates. Staff stability (standardized $\gamma = .43$, $p < .05$) and teacher professional quality (standardized $\gamma = .22$, $p < .05$) were also significantly related to school social/curricular organization but not to change in leadership or improvement capacity. Next, we examine the two research questions posed for this study.

What is the relationship between distributed leadership and academic capacity when observed over time? This question examined proposed relationships among variables in the model in terms of their initial levels and subsequent levels. The results provide support for our first proposition that the initial level of distributed leadership in the school would be related to subsequent change in academic capacity (standardized $\gamma = .14$, $p < .05$). Similarly, initial level of academic capacity was significantly related to subsequent change in distributed leadership (standardized $\gamma = .19$ $p < .05$). Coding of the growth factors in the model (0, 1, *) implies the growth intervals between Occasion 1 and Occasion 2 were fixed (representing linear change), while the estimated factor loadings for the third occasion incorporate possible nonlinear change. Therefore, the path coefficients for the change constructs should be interpreted as the amount of change occurring between Year 1 and Year 3 associated with a 1 standard deviation increase in the levels of the initial factors. This implies that a 1 standard deviation increase in initial leadership would yield a 0.14 standard deviation increase in the academic capacity growth rate between Years 1 and 3. Similarly, a 1 standard deviation increase in initial capacity would yield about a 0.19 standard deviation increase in the leadership growth rate between Years 1 and 3.

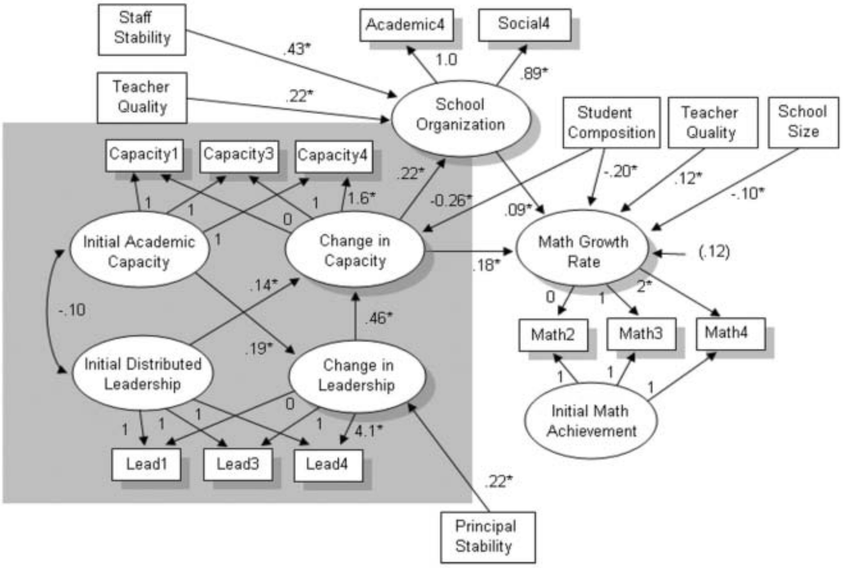

Figure 2. **Model 4 Between-School Standardized Effects**
$^*p < .05$.

How does distributed leadership impact school improvement capacity and subsequent growth in math? This question focused on the effects of changes in distributed leadership and capacity building (as perceived by teachers and students) and learning outcomes over the 4-year period. First, we proposed that change in distributed leadership would be *directly* and significantly related to change in schools' academic capacity. Since leadership is often seen as a catalyst for change, we hypothesized that stronger perceptions of leadership would be associated with increased academic capacity. As proposed, we found change in distributed leadership was moderately and significantly related to change in academic capacity (standardized $\gamma = .46$, $p < .05$). We also tested whether the hypothesized path might instead be in the other direction (i.e., from change in capacity to change in leadership) but found that Model 4 provided a superior fit to the data (see Model 5 in Table 1).

Second, we proposed that changes in academic capacity would be *directly* and significantly related to (a) sociocurricular organization and (b) growth in student learning. We found that change in academic capacity and student growth rates in math was also significant and substantial (standardized $\gamma = .18$, $p < .05$). We noted that this relationship was somewhat stronger (standardized $\gamma = .26$, not tabled) before sociocurricular organization was added to the model. Controlling for sociocurricular organization, then, was useful in estimating the size of the effect associated with changing academic capacity on student growth more accurately. We also found that teacher perceptions of changes in academic capacity were positively related to student perceptions of sociocurricular organization in their classrooms (standardized $\gamma = .22$, $p < .05$) and that sociocurricular organization was positively related to growth in math (standardized $\gamma = .09$, $p < .05$). Although not providing a direct test of our proposed model concerning how school efforts to increase academic capacity can result in specific classroom changes, these results provide some indirect support for that premise, in that teacher and student perceptions within schools corresponded positively.[7]

Our third proposition stated that the combined effects of distributed leadership on student growth rates in math would be indirect rather than direct. The indirect effects of change in distributed leadership (mediated by change in academic capacity) on sociocurricular organization (standardized $\gamma = .10$) and student growth rates (standardized $\gamma = .09$) were significant ($p < .05$, not tabled). Although the size of the indirect effect of distributed leadership on student growth in math may appear small, it is on a par with the direct effects of other variables in our model (i.e., teacher professional preparation and school sociocurricular organization) known from previous studies to affect learning outcomes (e.g., Betts, Rueben, & Danenberg, 2000).

Comparison of effect sizes among school-level variables may be a more accurate means of judging the size of school effects than simply adopting language such as *small* or *medium* to describe them (Bloom et al., 2008). More specifically, standardized effects of 0.1 would increase school growth rates by about 20% (0.10/0.50), and standardized effects of 0.2 would increase school growth rates by about 40% (see Bloom et al., 2008). As Table 1 suggests, Model 4 accounted for an additional 13% of the total variance in math growth beyond the Model 2 variables. The total variance accounted for in school growth in math was 88%, with 12% from other sources (in parentheses in Figure 2).

We also tested the validity of the proposed model against a model incorporating both direct and indirect leadership effects on math growth rates. This alternative model included one more parameter representing the direct path between change in distributed leadership and math growth (see Figure 2). The model with both indirect and direct effects (Model 6 in Table 1) did not provide an improvement in fit, compared against the model with only indirect effects (adjusted for nonnormality, $\Delta\chi^2 = 0.06$, $p > .05$) with $\Delta df = 1$. Moreover, the estimate of direct leadership effects on growth rates was not significant (standardized $\gamma = .01$, $p > .10$, not tabled). We therefore accepted the indirect leadership effects model (Model 4 in Table 1) as a more parsimonious representation. This result is relevant to the ongoing theoretical issue of whether leadership effects on school improvement outcomes should be conceptualized as indirect only or both direct and indirect.[8]

Finally, we proposed that selected context variables could moderate the exercise of school improvement leadership. This analysis examined whether the model of leadership effects on school improvement might vary systematically across different types of school settings. Figure 2 indicates only a few significant effects of context variables on change in distributed leadership and change in academic capacity. More specifically, in schools where the same principal was present over the period of the study, teachers reported greater capacity-building in distributed leadership over time (standardized $\gamma = .22$, $p < .05$). Student composition did not affect change in distributed leadership (standardized $\gamma = .06$, $p > .10$, not tabled). When we added an interaction of these two variables (Principal Stability × Composition), it was also not significantly related to perceptions of change in leadership (standardized $\gamma = -.05$, $p > .05$, not tabled). Although student composition was related to perceptions of change in academic capacity (standardized $\gamma = -.26$, $p < .05$), principal stability was not.

Discussion

This study built on a substantial body of research that has explored the effects of leadership on school improvement and student learning (Hallinger & Heck, 1996, 1998; Leithwood et al., 2004, in press; Robinson et al., 2008). Initially, we highlighted two limitations of this knowledge base that this study sought to address: a lack of empirical research on the effects of distributed leadership and longitudinal research concerning leadership effects on school improvement. In this section, we summarize the findings, review limitations of the research, and outline the implications.

First, we proposed that the relationship between distributed leadership and academic capacity was dynamic and possibly reciprocal. The limitations of viewing leadership solely as the causal factor in school improvement change have been amply discussed in the literature (e.g., Hallinger & Heck, 1996; Krüger et al., 2007; Luyten et al., 2005; Pitner, 1988; Witziers et al., 2003). In this study, we employed a multidimensional perspective that focused on several aspects of school organization hypothesized to influence growth in student learning. Since these constructs are not readily amenable to experimental manipulation, we relied on longitudinal panel data, in which the key constructs proposed to drive school improvement were measured on multiple occasions over a 4-year period (Cook, 2002; Marsh & Craven, 2006).

We found support for the hypothesis that school leadership and capacity building are mutually reinforcing in their effects on each other over time. This reciprocal effects model of school improvement is underpinned by the notion that in settings where people perceive stronger distributed leadership, schools appear better able to improve their academic capacity. Similarly, where academic capacity is perceived to be stronger at one point in time, this appears to be advantageous to the development of stronger leadership over time. Change in these organizational processes has been proposed to underpin school improvement (Fullan, 2006).

Second, we found changes in these mutually reinforcing constructs were also positively associated with school growth rates in math. The effect size for change in academic capacity was almost 0.2. This implies that an increase of 1 standard deviation in change in capacity would be associated with an increase in the average school growth rate of almost 40%. The finding of *indirect* leadership effects on learning growth rates extends an important conclusion from previous cross-sectional research (Bell et al., 2003; Hallinger & Heck, 1996; Leithwood et al., 2004; Robinson et al., 2008; Witziers et al., 2003).

The focus on distributed school leadership is of theoretical interest and practical importance. Up to now, the literature on distributed leadership has emphasized conceptual development (Gronn, 2002) and description of distributed leadership practices (Leithwood et al., 2009; Spillane, 2006). Our findings represent an early contribution to the emerging empirical knowledge base on the effects of distributed school leadership (e.g., Marks & Printy, 2003; Mulford & Silins, 2003; Pounder et al., 1995). The study highlights additional sources of school leadership and explicitly links distributed leadership to capacity-building strategies designed to impact teaching and learning.

Our findings imply the need to distribute particular types of leadership practices and create a sustained focus on strategies aimed at the improvement of teaching and learning (e.g., fostering curricular standards and alignment, developing instruction, providing tangible support for students, improving professional capacity, sustaining a focus on academic improvement). Unfortunately, given limitations in measurement of the leadership construct, our results offer little direct insight into which leadership practices should be distributed or how they should be distributed among different staff roles.

Third, the results also suggested that changes in teacher perceptions of distributed leadership and academic capacity were significantly related to student perceptions of the quality of the school's sociocurricular organization. This relationship also supports the validity of our proposed school improvement model because the evidence comes from different sources. Moreover, even after adding this additional mediating variable to our proposed improvement model, both leadership and academic capacity effects remained significantly related to math growth rates (although slightly diminished).

Finally, although this study did not explicitly measure the contribution of principal leadership to building academic capacity, principal stability demonstrated a small, but statistically significant, positive effect on teacher perceptions of changes in distributed leadership. In schools where the same principal was present over the course of the study, there was a significantly stronger perception of academic capacity at the end of the 4 years. One possible interpretation of this finding is that successful principals tend to stay longer at their schools. Another is that the principal's leadership role may remain important even when schools are seeking to develop a broader capacity for leadership. As some theorists have speculated, supportive leadership from the principal may well represent a necessary (but not sufficient) condition to developing the capacity among other school leaders (Barth, 1990, 2001; Fullan, 2001; Leithwood et al., 2009).

Limitations

These findings should be interpreted in light of several limitations. First, caution must be exercised in the use of SEM applications to test substantive theories in nonexperimental designs. Uncontrolled (omitted) and confounding variables are a common source of misspecification that can produce misleading results. Our proposed model focused primarily on the mediating effects of distributed leadership, academic capacity, and sociocurricular organization on growth rates in math achievement. Student growth, however, is determined by other school and classroom variables as well (Creemers, 1994; Darling Hammond, 2006). There may be other mediators at work, and these may also be correlated with leadership and capacity building.

Such variables could include grouping strategies used in assigning students to classrooms (Burns & Mason, 1998) or teacher effectiveness (McCaffrey et al., 2004). In this state database, individual teacher classroom effectiveness and collective school teaching effectiveness account for substantial variance in student math outcomes (Heck, 2009), but we were unable to link this particular student cohort to their specific teachers over time. Further research may add classroom data as a third level to the analysis.

Moreover, although the longitudinal analyses revealed evidence of change in model constructs over time, they do not provide complete protection against a selection-bias argument. For example, teachers may perceive improvement capacity or distributed leadership more positively in schools that achieve at high levels over longer periods of time than the 4 years of this study. Even though we controlled for initial achievement level in our model, the achievement contexts of schools (and their unknown effects on variables) represent a possible confounding variable. In longitudinal panel studies, attrition (e.g., staff mobility) also represents a possible confounding variable (Robinson & Marsland, 2008). Although teacher turnover rates were relatively modest each year (about 8%), it remains unknown how this might affect schoolwide measures of change.

Second, questions remain about the definition and measurement of distributed leadership and academic capacity as collective properties of schools. Measurement error can contribute to model misspecification, which can produce misleading results (Bentler & Bonett, 1980). We took preliminary steps to assess possible changes in psychometric properties associated with measuring the constructs on multiple occasions (Collins, Cliff, & Dent, 1988). We found the measurement properties of our constructs to be reliable. Despite this, annual school-level questionnaires are admittedly imperfect means of extracting information about organizational processes. For example, an individual's reported involvement in school decision making may, or may not, adequately capture a key aspect of distributed leadership, and even if it does, the way the individual's reply is coded into a score may bias its exact meaning (Bentler & Bonett, 1980).

Although we found that changes in levels of distributed leadership were related to changes in academic capacity and indirectly to student growth, questions remain concerning the nature of day-to-day implementation of leadership efforts aimed at improving academic capacity. Academic capacity is only a proxy for more thorough information that could be assembled about schoolwide efforts to improve curriculum and teacher expertise, as well as teacher instructional behavior in classrooms (Cohen & Hill, 2000; Creemers, 1994). School-level aggregates ignore wide variations in teaching and learning conditions that may be important at the classroom level (McCaffrey et al., 2004).

Finally, questions also remain about the temporal sequence underlying associations between distributed leadership, academic capacity, and growth in student learning. Constructing a proper temporal sequence remains a consistent limitation of previous studies looking at the relationship between school leadership and school processes. Although the 4-year period of this study provides additional leverage over the limitations over cross-sectional studies, it might require an even longer time frame in order to observe patterns of change in some organizational processes. Thus, we note that it remains a challenge to disentangle temporal effects in organizational studies, since one must always "jump" into a temporal sequence at some arbitrary point in time. Although the study begins to address the issue of temporal relationships, further research is needed to refine proposed causal relationships and to eliminate possible rival explanations.

Implications

Despite these limitations, our results have several implications for research, policy, and practice. First, the research demonstrates the utility of longitudinal panel studies for modeling simultaneous change among several sets of organizational variables. We believe that this represents a useful foundation for future research on leadership effects, since school improvement, by definition, involves change over time.

Second, publication of several influential reviews of research in the 1980s (Bossert et al., 1982; Bridges, 1982; Leithwood & Montgomery, 1982; Pitner, 1988) gave impetus to the more systematic empirical study of school leadership and its effects (Hallinger & Heck, 1996). Although progress has been made at identifying and specifying the nature of principal leadership effects (Bell et al., 2003; Hallinger & Heck, 1996, 1998; Leithwood et al., 2004, in press; Robinson et al., 2008; Witziers et al., 2003), it is also true that the powerful effects attributed to school leadership by policy makers have yet to be fully validated through research (Heck & Hallinger, 2005).

Our study suggests that a historically narrow focus on the impact of principal leadership may have hid a portion of the school's leadership resources from our conceptual and empirical lenses. We would note that the indirect effects of distributed leadership on student learning found in this study were larger than found in many of the cross-sectional studies (e.g., Hallinger et al., 1996; Heck et al., 1990; Wiley, 2001). Whether the difference in magnitude of indirect effect was due to differences in our conceptualization of leadership as an organizational property rather than an individual attribute of the principal, or due to differences in the research design (i.e., cross-sectional vs. longitudinal), is an issue on which we cannot speculate at this time.

Over the past decade, emergent recognition of the boundaries of what principals can accomplish in the practical world of schools has led scholars to evince greater interest in conceptualizations of distributed leadership (Gronn, 2002; Spillane, 2006). Our findings with respect to the modeling of distributed leadership support the ongoing validation of this construct and offer insight into its relationship to other key improvement factors. Future research on school leadership will likely benefit by incorporating an explicit measure of principal leadership as well as a broader measure of shared leadership from other sources.

Third, with respect to policy, our research focuses attention on a set of key organizational processes (i.e., distributed leadership, academic capacity) that may be linked to successful school improvement. Distributed leadership appeared to contribute to the development of academic capacity and indirectly to student learning outcomes. Thus, the findings provide empirical support for calls for the development of broader and deeper capacity to lead in schools (Barth, 1990, 2001; Fullan, 2001; Lambert, 2002).

Our results add to the incremental process of knowledge building in the domain of school leadership effects. Validation of these findings will require researchers to follow schools for longer periods of time and conduct analyses that link changes in leadership and school organization with changes in teacher practices and student learning. Nevertheless, we conclude that these empirical results strongly support the continuation of this line of longitudinal inquiry into school leadership effects, which, heretofore, has only been supported in conceptual analyses.

Notes

The authors are grateful for helpful comments received from Ed Bridges, Ellen Goldring, Ken Leithwood, George Marcoulides, Bill Mulford, Viviane Robinson, and two anonymous reviewers on earlier versions of the article.

[1] As is common in the school effectiveness literature, we use the term *school effects* to indicate statistically significant associations between variables. These associations do not need to be causal in nature.

[2] During the years of the study, of approximately 3,000 new teachers hired, 43% had completed a teacher education program but had not finished all PRAXIS examinations when hired, and 22% had not completed a teacher education program (Department of Education, 2006).

[3] We also used similar student perceptions ($\alpha = .95$) of standards implementation and learning in their classrooms during the last year of the study as an additional measure to triangulate teachers' perceptions of classroom changes.

[4] One possible reason for the lack of association between school context and teacher mobility is that mobility is driven primarily by retirement patterns and the union contract specifying transfer and hiring procedures (based on seniority) rather than by principal discretion in hiring teachers.

[5] Matrices and vectors facilitate the specification of latent change analysis models. For math, the model to represent individual i at time t can be written as

$$y_{it} = v_t + \Lambda_t \eta_i + K x_i + \varepsilon_{it}, \tag{1}$$

where y_{it} is a vector of math outcomes for individual i at time t ($y_{i1}, y_{i2}, \ldots, y_{iT}$), v_t is a vector of measurement intercepts, Λ_t is a $p \times m$ design matrix representing the change process, η_i is an n-dimensional vector of latent variables ($\eta_{0i}, \eta_{1i} \ldots, \eta_{pi}$), K is a $p \times q$ parameter matrix of regression slopes relating x_i covariates ($x_{1i}, x_{2i} \ldots, x_{pi}$) to the latent factors, and ε_{it} represents time-specific errors that are contained in a covariance matrix (Θ). The factor loadings for the latent initial status and change latent variables are defined in the Λ_t factor loading matrix. After measuring the latent factors, the second part of the analysis concerns the structural relationships between the latent variables and other covariates. Variability in initial math levels (η_{0i}) and change (η_{1i}) can be modeled as a function of one or more covariates (x_i) plus error:

$$\eta_{0i} = \alpha_0 + \gamma_0 x_i + \zeta_{0i}, \tag{2}$$

$$\eta_{1i} = \alpha_1 + \gamma_1 x_i + \zeta_{1i}, \tag{3}$$

where α_0 and α_1 are measurement intercepts and γ_0 and γ_1 are structural parameters describing the regressions of latent variables on a covariate. Each latent factor has its own residual (ζ_{0i}, ζ_{1i}) that permits the quality of measurement associated with each individual's growth trajectory to differ from those of other individuals. The individual growth model can be further divided into respective individual-level and school-level components (see Muthén & Muthén, 1998–2006). The leadership and academic capacity change processes were defined in a similar manner but only at the school level.

[6] Standardized estimates were .02 (female), −.05 (low SES), .13 (English language learner), −.05 (special education), −.11 (minority), and −.08 (changed schools).

[7] For example, we found their perceptions to be substantially correlated on the Year 4 standards emphasis and implementation indicator of academic capacity ($r = .64, p < .01$).

[8] Most recent commentary in the field tends to emphasize the validity of an indirect effects model (e.g., Leithwood et al., 2004; Robinson et al., 2008; Southworth, 2002). Nonetheless, we note that in a discussion of this article at the annual meeting of the American Educational Research Association in 2009, the discussant, Professor Brian Rowan, asserted that the field should not entirely discount the possibility that direct effects of principal leadership may also be at work.

References

Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review*, *47*(5), 626–640.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Barth, R. (1990). *Improving schools from within*. San Francisco: Jossey-Bass.

Barth, R. (2001). Teacher leader. *Phi Delta Kappan*, *82*(6), 443–449.

Bell, L., Bolam, R., & Cubillo, L. (2003). *A systematic review of the impact of school head teachers and principals on student outcomes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Bentler, P., & Bonett, D. (1980). Significance of tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California*. San Francisco: Public Policy Institute of California.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328.

Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34–64.

Braddock, J. H., & Slavin, R. E. (1993). Why ability grouping must end: Achieving excellence and equity in American education. *Journal of Intergroup Relations*, *20*, 51–64.

Bridges, E. (1982). Research on the school administrator: The state-of-the-art, 1967–1980. *Educational Administration Quarterly*, *18*(3), 12–33.

Burns, R. B., & Mason, D. A. (1998). Class formulation and composition in elementary schools. *American Educational Research Journal*, *35*(4), 181–207.

Calsyn, R. J., Winter, J. P., & Burger, G. K. (2005). The relationship between social anxiety and social support in adolescents: A test of competing causal models. *Adolescence*. *40*(157), 103–113.

Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Cohen, D., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, *102*(2), 294–343.

Collins, L. M., Cliff, N., & Dent, C. (1988). The longitudinal Guttman simplex: A new methodology for measurement of dynamic constructs in longitudinal panel studies. *Applied Psychological Measurement*, *12*(3), 217–230.

Cook, T. (2002). Randomized experiments in education. Why are they so rare? E*ducational Evaluation and Policy Analysis*, *24*(3), 175–200.

Crandall, D., Eiseman, J., & Louis, K. S. (1986). Strategic planning issues that bear on the success of school improvement efforts. *Educational Administration Quarterly*, *22*(3), 21–53.

Creemers, B. P. (1994). *The effective classroom*. London: Cassell.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Darling Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational Researcher*, *35*(7), 13–24.

Davies, R. (1994). From cross-sectional to longitudinal analysis. In R. Davies & A. Dale (Eds.), *Analyzing social & political change: A casebook of methods* (pp. 20–40). Newbury Park, CA: Sage.

Day, D., Gronn, P., & Salas, S. (2006). Leadership in team-based organizations: On the threshold of a new era. *The Leadership Quarterly*, *17*(3), 211–216.

Department of Education. (2006). *Teacher employment report, 2005–2006*. Honolulu, Hawaii: Author.

Donaldson, G. A. (2001). *Cultivating leadership in schools: Connecting people, purpose, and practice*. New York: Teachers College Press.

Firestone, W. A., & Wilson, B. L. (1985). Using bureaucratic and cultural linkages to improve instruction: The principal's contribution. *Educational Administration Quarterly*, *20*(2), 7–30.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.

Fullan, M. (2006). *The development of transformational leaders for educational decentralization*. Toronto, Canada: Michael Fullan.

Geijsel, F., Sleegers, P., Stoel, R., & Krüger, M. (2009). The effect of teacher psychological, school organizational and leadership factors on teachers' professional learning in Dutch schools. *Elementary School Journal*, *109*(4), 406–427.

Goldhaber, D. D. (2002). The mystery of good teaching. *Education Next*, *2*(1), 50–55.

Gronn, P. (2002). Distributed leadership as a unit of analysis. *The Leadership Quarterly*, *13*, 423–451.

Gross, N., & Herriot, R. (1965). *Staff leadership in schools*. New York: Wiley.

Hall, G., & Hord, S. (2001). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn & Bacon.

Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership and student achievement. *Elementary School Journal, 96*(5), 498–518.

Hallinger, P., & Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of the empirical research, 1980–1995. *Educational Administration Quarterly*, *32*(1), 5–44.

Hallinger, P., & Heck, R. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, *9*(2), 157–191.

Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education*, *94*(3), 328–355.

Harris, A. (2003). Teacher leadership as distributed leadership: Heresy, fantasy or possibility? *School Leadership and Management*, *23*(3), 313–324.

Heck, R. (2009, April). *Examining the impact of successive teachers on student learning: Testing a multilevel cross-classified model*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, California.

Heck, R., & Hallinger, P. (2005). The study of educational leadership and management: Where does the field stand today? *Educational Management, Administration & Leadership*, *33*(2), 229–244.

Heck, R., Larsen, T., & Marcoulides, G. (1990). Principal instructional leadership and school achievement: Validation of a causal model. *Educational Administration Quarterly*, *26*, 94–125.

Hill, P., & Rowe, K. (1996). Multilevel modeling in school effectiveness research. *School Effectiveness and School Improvement*, *7*(1), 1–34.

Hoyle, R., & Panter, A. (1995). Writing about structural equation models. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Newbury Park, CA: Sage.

Jackson, D. (2000). The school improvement journey: Perspectives on leadership. *School_Leadership & Management*, *20*(1), 61–78.

Jöreskog, K., & Sörbom, D. (1989). *LISREL user's guide*. Chicago: Scientific Software.

Krüger, M., Witziers, B., & Sleegers, P. (2007). The impact of school leadership on school level factors: Validation of a causal model. *School Effectiveness and School Improvement*, *18*(1), 1–20.

Lambert, L. (2002). A framework for shared leadership. *Educational Leadership, 59*(8), 37–40.

Lee, V., & Bryk, A. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*(3), 172–192.

Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal*, *40*(3), 353–393.

Leithwood, K., Anderson, S., Mascall, B., & Strauss, T. (In press). School leaders' influences on student learning: The four paths. In T. Bush, L. Bell, & D. Middlewood (Eds.), *The principles of educational leadership and management*. London: Sage.

Leithwood, K., & Jantzi, D. (1999). The relative effects of principal and teacher leadership on student engagement in school. *Educational Administration Quarterly, 35*(Suppl.), 679–706.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlsttom, K. (2004). *Review of research: How leadership influences student learning*. Wallace Foundation. Retrieved December 19, 2007, from http://www.wallacefoundation.org/NR/rdonlyres/E3BCCFA5-A88B-45D3-8E27-B973732283C9/0/ ReviewofResearchLearningFromLeadership.pdf.

Leithwood, K., Mascall, B., & Strauss, T. (2009). What we have learned, where we go from here. In K. Leithwood, B. Mascall, & T. Strauss (Eds.), *Distributed leadership according to the evidence* (pp. 269–282). New York: Routledge.

Leithwood, K., & Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research*, *52*(3), 309–339.

Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, *16*(3), 249–279.

Marks, H., & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370–397.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67–101.

Mulford, B., & Silins, H. (2003). Leadership for organisational learning and improved student outcomes—What do we know? *Cambridge Journal of Education*, *33*(2), 175–195.

Muthén, L., & Muthén, B. (1998–2006). *Mplus user's guide*. Los Angeles: Author.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.

Ogawa, R., & Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, *31*(2), 224–243.

Peugh, J. L., & Enders, C. K. (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(1), 1–19.

Pitner, N. (1988). The study of administrator effects and effectiveness. In N. Boyan (Ed.), *Handbook of research in educational administration* (pp. 99–122). New York: Longman.

Pounder, D. G., Ogawa, R. T., & Adams, E. A. (1995). Leadership as an organization-wide phenomena: Its impact on school performance. *Educational Administration Quarterly*, *31*(4), 564–588.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Reynolds, D., Teddlie, C., Hopkins, D., & Stringfield, S. (2000). Linking school effectiveness and school improvement. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 206–231). London: Falmer.

Robinson, S., & Marsland, L. (2008). Approaches to the problem of respondent attrition in a longitudinal panel study of nurses' careers. *Journal of Advanced Nursing*, *20*(4), 729–741.

Robinson, V., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 564–588.

Selzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress. Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis*, *25*(3), 263–286.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Sleegers, P., Geijsel, F., & Van den Berg, R. (2002). Conditions fostering educational change. In K. Leithwood (Ed.), *Second international handbook of educational leadership and administration* (pp. 75–102). Dordrecht, the Netherlands: Kluwer Academic.

Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership and Management*, *22*(1), 73–92.

Southworth, G. (2003). *Primary school leadership in context: Leading small, medium and large sized schools*. London: Taylor and Francis.

Spillane, J. (2006). *Distributed leadership*. San Francisco: Jossey-Bass.

Stoll, L., & Fink, D. (1996). *Changing our schools*. London: Open University Press.

Teddlie, C., Stringfield, S., & Reynolds, D. (2000). Context issues within school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *International handbook of school effectiveness research* (pp. 160–183). London: Routledge.

Wiley, S. (2001). Contextual effects on student achievement: School leadership and professional community. *Journal of Educational Change*, *2*(1), 1–33.

Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*(3), 398–425.

York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research*, *74*(3), 255–316.

Manuscript received September 11, 2008
Final revision received May 20, 2009
Accepted May 20, 2009