The Leadership Quarterly 21 (2010) 867–885

Contents lists available at ScienceDirect

The Leadership Quarterly

journal homepage: www.elsevier.com/locate/leaqua

Testing a longitudinal model of distributed leadership effects on school improvement

Ronald H. Heck [a,*], Philip Hallinger [b]

[a] Educational Administration and Policy, College of Education, University of Hawaii at Manoa, 1776 University Avenue, Honolulu, Hawaii 96822, United States
[b] Hong Kong Institute of Education, Hong Kong

ARTICLE INFO

Keywords:
Distributed leadership
School leadership
School improvement
Organizational change

ABSTRACT

A central premise in the literature on leadership highlights its central role in organizational change. In light of the strength of this conceptual association, it is striking to note the paucity of large-scale empirical studies that have investigated how leadership impacts performance improvement in organizations over time. Indeed evidence-based conclusions concerning the impact of leadership on organizational change are drawn largely from case studies and cross-sectional surveys. Neither approach satisfies the design requirements for studying the contribution of leadership to performance improvement in organizations. This paper tests a longitudinal, multilevel model of change in distributed leadership, school improvement capacity, and student performance over a four-year period. The results suggest that change in distributed leadership and organizational capacity for improvement make significant contributions to growth in student learning in reading and math.

Scholars have conducted extensive inquiry into the nature of organizational leadership over the past several decades (Bass, 1985, 1990; Yukl, 2006). One fruitful line of inquiry has employed "mediated-effects" models (Hallinger & Heck, 1996a; Pitner, 1988) to explicate and test conceptual linkages between leadership, organizational structures and processes, and performance outcomes (Bass, 1985, 1990; Bass & Avolio, 1994; Campbell, McCloy, Oppler, & Sager, 1993; Lord, 2001; Mohr, 1983; Mumford, Zaccaro, Johnson, Diana, Gilbert, & Threlfall, 2000; Steers, 1975; Thomas, 1988). Using a similar approach, scholars studying leadership in educational organizations have found positive, indirect effects of principal leadership on student learning (Bell, Bolam, & Cubillo, 2003; Hallinger & Heck, 1996a; Leithwood, Louis, Anderson, & Wahlsttom, 2004; Witziers, Bosker, & Kruger, 2003). Although the effect size in these studies tends to be small, researchers have suggested that the level of impact is meaningful because the effects of better schooling (e.g., curriculum, academic expectations, teaching and leadership effectiveness) are hypothesized to accumulate during students' tenure at a particular school (Hallinger & Heck, 1998; Leithwood, Day, Sammons, Harris, & Hopkins, 2006; Leithwood et al., 2004; Robinson, 2007).

The same level of substantive progress is, however, less evident when attention turns to the relationship between leadership and improvements in organizational performance (Monge, 1990; Nonaka & Toyama, 2002). Although scholars place leadership at the center of efforts to bring about change in organizations (e.g., Bass, 1985; Bennis, 2003; O'Toole, 1995; Yukl, 2006), only recently has empirical research begun to explore the means by which leadership contributes to *change* in performance outcomes (Atwater, Dionne, Avolio, Camobreco, & Lau, 1999; Edmondson, Roberto, & Watkins, 2003; Glover, Rainwater, Friedman, & Jones, 2002; Hooijberg & Schnieder, 2001). Similarly, despite a thriving literature on leadership effects in education, there are remarkably few large-scale empirical studies of how leadership impacts *improvement* in the academic performance of students and their schools *over time* (Heck & Hallinger, 2005; Leithwood et al., 2004, 2006; Reynolds, Teddlie, Hopkins, & Stringfield, 2000; Sleegers,

* Corresponding author.
E-mail address: rheck@hawaii.edu (R.H. Heck).

1048-9843/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.leaqua.2010.07.013

Geijsel, & Van den Berg, 2002). Indeed, most relevant commentary on this issue is based on personal experience, case studies (e.g., Foster, 2005; Jackson, 2000; Stoll & Fink, 1996) and cross-sectional surveys of school (e.g., Edmonds, 1979; Hill & Rowe, 1996) and leadership effectiveness (e.g., Edmonds, 1979; Hallinger et al., 1996; Heck et al., 1990; Wiley, 2001). Unfortunately, neither approach meets the requirements necessary for drawing evidence-based conclusions about the relationship between leadership and school improvement.

This paper presents findings from a study that examined the effects of leadership on school improvement. The study was conducted within the context of federal and state efforts to improve school accountability and performance in the United States [e.g., No Child Left Behind (NCLB), 2001]. In recent years, educational policymakers in diverse nations (e.g., USA, Australia, New Zealand, Hong Kong, United Kingdom) have targeted leadership, in many cases "distributed" (or shared) leadership (Gronn, 2002, 2003; Spillane, 2006), as a means of building a more productive, learning-focused organizational climate in schools (Harris, 2003; Leithwood, Mascall, & Strauss, 2009; Leithwood et al., 2004, 2006) In the state where this particular study took place, the legislature had recently granted increased authority to elected school councils as well as to teachers to exercise leadership directed toward school improvement.

In this report, we examine the effects of *changes* in distributed leadership on *changes* in school improvement capacity and *growth* in student learning. More specifically, we present findings from the development and testing of a longitudinal, multilevel model (Hill & Rowe, 1996; Raudenbusch & Willms, 1995) that estimated changes in the relationships among these variables over a four-year period. Our study offers two main contributions to leadership research.

First, the study adds to an emerging empirical literature on distributed leadership in schools. Despite calls for studies that test policy prescriptions for distributed leadership against empirical evidence, most research to date has focused on the description of distributed leadership practices in schools (Leithwood et al., 2009). Few large-scale studies have examined the impact of distributed leadership on people, processes, or performance outcomes in schools (Harris, 2003; Leithwood et al., 2006, 2009; Timperly, 2009).

Second, this research addresses a persisting question of interest in the leadership effects literature by examining whether the small indirect effects of school leadership on student learning observed in cross-sectional studies become more substantial as they unfold over time (Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger & Heck, 1996b; Kaiser, Hogan, & Craig, 2008; Leithwood et al., 2004; Marcoulides & Heck, 1993; Robinson, 2007). The study explored whether the application of structural equation modeling (SEM) techniques with longitudinal data might offer a more robust approach for illuminating the nature of these relationships than traditional cross-sectional analyses of leadership effectiveness (Heck & Thomas, 2009; Huber & Van de Ven, 1995; Mathieu et al., 2007; Podsakoff, 1994; Raykov & Marcoulides, 2006).

1. Defining a longitudinal model of school improvement leadership

In this section we begin by delineating the rationale for using longitudinal designs in research on leadership and change. We then present the multilevel model of leadership and change used in this study. This is followed by the conceptual definition of variables in the model.

1.1. Rationale for longitudinal studies of leadership effects

A prominent line of inquiry in leadership research has focused on understanding the contributions that leadership makes to organizational processes and performance outcomes (Bass & Avolio, 1994; Gresov, Haveman, & Oliva, 1993; Langlois & Robertson, 1993; Mathieu, Ahearne, & Taylor, 2007; Nissen & Levitt, 2002; Sivasubramaniam, Murry, Avolio, & Jung, 2002; Steers, 1975; Tate, 2008; Teece, 1982; Tushman & Romanelli, 1985). A common approach has been to collect survey data at one point in time that describe the behaviors and activities across a sample of leaders and analyze relationships with selected organizational variables and measures of performance or effectiveness (Bass, 1990; Campbell et al., 1993; Hallinger & Heck, 1996a; Kaiser et al., 2008). The same approach cannot, however, be employed when the question turns to leadership and *organizational change and improvement*.

Theories that seek to explain change in social phenomena typically focus on two or more variables that are proposed to change concomitantly over time (Blalock, 1989). Yet, temporal relationships between variables (e.g., leadership and staff performance) cannot be adequately modeled through data collected at a single point in time (Blalock, 1989; Campbell & Stanley, 1966; Davies, 1994; Heck & Hallinger, 2005; Kerlinger, 1986; Podsakoff, 1994). Ogawa and Bossert (1995) state the case for using longitudinal designs in this domain of research:

> [S]tudies of leadership must have as their unit of analysis the organization. Data on the network of interactions that occur in organizations must be compiled over time….The importance of the dimension of time must be emphasized. If leadership involves influencing organizational structures, then time is important. Only time will tell if attempts at leadership affect organizational solidarity. Also, the time that is required for such effects to occur and the duration of the persistence of the effects may be important variables. (pp. 239–240)

This suggests that modeling leadership effects on organizational change requires the collection and analysis of longitudinal data. However, the conduct of longitudinal studies on a scale sufficient to assess the impact of leaders across organizations poses resource, logistical, and technical challenges for researchers (Heck & Hallinger, 2005; Kelly & McGrath, 1988; Podsakoff, 1994; Singer & Willett, 2003; Willms, 1992). In particular, scholars have highlighted the stringent data requirements as well as the need for analytic techniques with the capability to model change among multiple variables over time across organizational levels (Kelly & McGrath, 1988; Nissen & Levitt, 2002; Podsakoff, 1994; Singer & Willett, 2003). Nonetheless, the constructs of leadership and organizational change are so intimately linked that there is little choice but to persist in the search for conceptual models and research methods that offer leverage for understanding their relationship (Glover et al., 2002; Holmqvist, 2003; Mathieu et al., 2007).

Fortunately, progress is evident. In recent years scholars have been aided by advances in analytical methods such as SEM that make it possible to incorporate temporal elements into empirical tests (Carroll & Burton, 2000; Davies, 1994; Heck & Thomas, 2009; Huber & Van de Ven, 1995; Mathieu et al., 2007; Ployhart, Holtz, & Bliese, 2002; Raykov & Marcoulides, 2006). Using these approaches, researchers have found that analyses of growth trajectories provide a stronger basis for making inferences about organizational performance than static measures. This is particularly true in situations where temporal sequences in organizational relationships are an integral aspect of the proposed model (Ployhart et al., 2002; Podsakoff, 1994; Willms, 1992).

1.2. Proposed model of leadership and change

Increasingly, educational systems throughout the world are holding the leadership of primary and secondary schools accountable for student performance results. Not surprisingly, and despite acknowledged measurement limitations, student achievement has become the key performance indicator favored by education policymakers from Hong Kong to Sydney and New York to London. Given the centrality of student achievement in national accountability systems and recent investments in the development of learning-centered leadership, this study selected student achievement as the key performance outcome measure for schools.

Recognition of the multilevel, nested structure of school organizations has been central to the development of the knowledge base on factors that impact student learning (Hill & Rowe, 1996; Raudenbusch & Willms, 1995; Teddlie & Reynolds, 2000). For example, Rutter noted (1983, p. 34), "[The] evidence...strongly suggests that it is meaningful to speak of the ethos of the school as a whole (while still recognizing marked variations between teachers and classrooms within any school)." Thus, when examining performance *between schools*, it is meaningful to note that students and classrooms are located in schools that possess varied social and educational capital (Howard, McLaughlin, & Vacha, 1996; Mulford, 2007). *Within schools*, students are clustered in classrooms with teachers who differ in terms of their preparation, skills, and academic expectations. Because the data structures that describe students, teachers, schools and their communities are nested, individuals will share similarities that must be considered in the selection and application of analytic methods (Lee & Bryk, 1989; Willms, 1992).

In this study we employed multilevel latent change analysis (LCA) to examine leadership and school improvement over a four-year period during which the state government implemented new educational policies aimed at improving school performance. In the LCA approach, changes in individual or organizational processes can each be represented by latent (or underlying) factors. A *level* factor represents the level of a particular variable at a chosen point in time. In this case, the level factors in our model describe the initial status of distributed leadership, school improvement capacity, and student learning in each school. A *shape* (or rate of change) factor represents change in the variable over a particular interval from these initial levels.

In this approach we assume, for example, that the capacity-building trajectories of individual schools (or the learning trajectories of individual students) have common algebraic forms, but that not every school (or student) has the same trajectory (Singer & Willett, 2003). In our study of latent change in schools, both the initial status (i.e., level factor) of students' achievement and their rates of learning growth (i.e., shape factor) are proposed to vary across schools. The subsequent focus of the research is to explain the school-level variability in student academic performance through sets of static and dynamic organizational variables.

Fig. 1 presents our proposed multilevel, longitudinal model of how school context, distributed leadership, and school improvement capacity are related to student learning outcomes. Our analytic approach facilitates the representation of both *static* and *dynamic* latent components associated with the leadership, school improvement capacity, and student learning constructs in the model. Because the constructs are treated as underlying processes that consist of correlated level and shape factors, each is represented with two ovals in the figure. Level (or static) components are proposed to affect *initial* student achievement. The shape (or dynamic) components reflect *change rates* in leadership and school improvement capacity which are proposed to affect student *growth rates*. Two-headed arrows between the level and shape factors indicate expected negative correlations between initial status and change in the proposed variables. The model proposes that, on average, schools with higher initial levels of distributed leadership, improvement capacity, and student learning will change in smaller increments over time (and the converse).

Our model also proposes that school leadership is influenced by the environmental (e.g., community socioeconomic status) and organizational context (e.g., school size, staff stability) in which it is exercised (Hallinger & Murphy, 1986; Leithwood et al., 2006). In the model, initial leadership is *indirectly* associated with initial student learning outcomes in math and reading. Change in levels of leadership over time is proposed to result in increased capacity to bring about school improvement which will carry over into enhanced growth in student performance.

1.3. Explanatory variables

As indicated in Fig. 1, the conceptual model incorporates three sets of explanatory variables: school context, distributed leadership, and school improvement capacity. Studies of school effectiveness (Edmonds, 1979; Hawley & Rosenholtz, 1984; Hill & Rowe, 1996; Mortimore, 1993; Rutter, 1983; Sammons, Nuttall, Cuttance, & Thomas, 1995), school improvement (Foster, 2005;

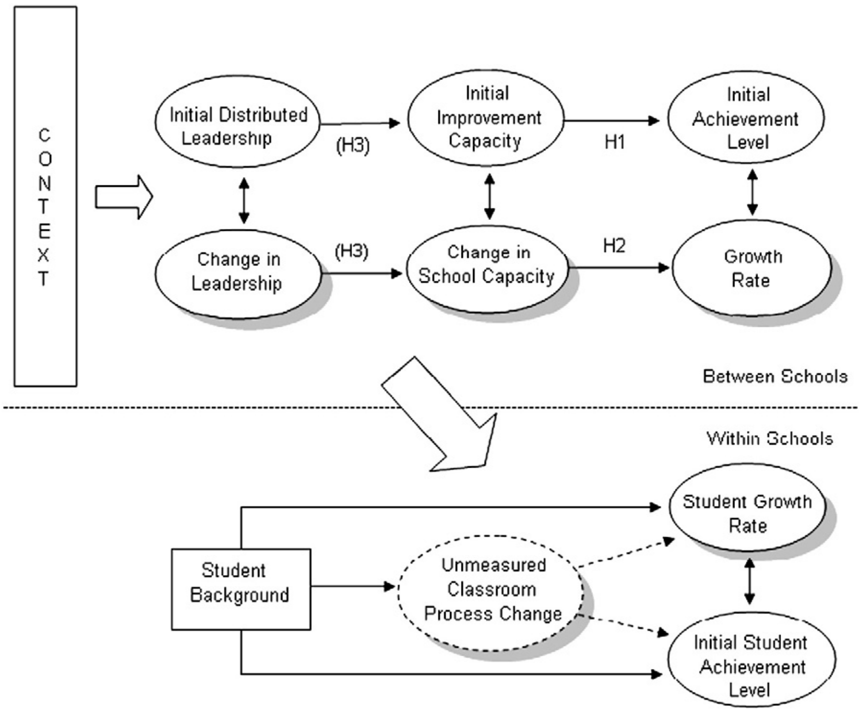

Fig. 1. Conceptual model of leadership and school improvement.

Jackson, 2000; Nicolaidou & Ainscow, 2005), effective classrooms (Creemers, 1994), and school leadership (Hallinger et al., 1996; Heck et al., 1990; Marks & Printy, 2003; Wiley, 2001) have contributed towards a more integrated understanding of how schools make a difference in learning (Ouston, 1999; Reynolds et al., 2000). This study incorporated key variables actively targeted in current school improvement efforts in its model of leadership and student learning.

1.3.1. Background and context variables

Early research on school effects identified the inequitable distribution of student learning resulting from student socio-economic background within and between schools (e.g., Coleman et al., 1966; Lee & Bryk, 1989). Students' social backgrounds influence grouping strategies as well as access to curriculum and quality teaching (Darling Hammond, 2000; Gamoran, 1986). Research further confirmed that interactions between the school's environment and its internal organization form a context in which school leadership is exercised (e.g., Bossert et al., 1982; Hallinger & Murphy, 1986; Leithwood et al., 2004, 2006; Ogawa & Bossert, 1995; Teddlie, Stringfield, & Reynolds, 2000). Background and context variables incorporated into this study included school size, teaching staff stability, principal stability, and student composition.

Consistent with previous research on schools, in Fig. 1 the broad arrow extending from school context indicates that these variables are proposed to affect school outcomes. The arrow also indicates that the school context may affect leadership and school improvement capacity. Although we acknowledge student composition as a contextual variable that influences many school processes, we do not specifically theorize about its specific effects in this report. It is included, however, as a control variable.

1.3.2. Distributed leadership

In Fig. 1, we include *distributed leadership* as a latent construct that is proposed to drive development of the school's capacity for improvement. Four assumptions frame the study's approach to leadership. First, the practice of leadership involves developing a vision for change and then motivating and enabling people to achieve the vision (Bass, 1990; Bennis, 2003; Leithwood et al., 2006; Yukl, 2006). Second, leadership in schools tends to be distributed and, therefore, its measurement should not be limited to the actions of those in formal management roles (Day et al., 2006; Gronn, 2002, 2009; Leithwood et al., 2006, 2009). Third, effective school leadership creates conditions that support teaching and learning and builds capacity for professional learning and change (Fullan, 2006; Hallinger et al., 1996; Hayes, Christie, Mills, & Lingard, 2004; Heck et al., 1990; Leithwood et al., 2004, 2006; Marks & Printy, 2003; Robinson, 2007; Wiley, 2001). Fourth, leadership that increases the school's capacity for improvement will impact student achievement positively (Bell et al., 2003; Fullan, 2006; Lee & Bryk, 1989; Lee, Croninger, & Smith, 1997; Leithwood et al., 2004; Marzano, Waters, & McNulty, 2005; Mulford & Silins, 2003; Robinson, 2007; Stoll & Fink, 1996).

Given its centrality to this study, the second assumption concerning distributed leadership requires additional elaboration. Although researchers have traditionally emphasized leadership exercised by those holding hierarchical positions, scholars have become increasingly interested in conceptions that highlight the distribution of leadership among individuals holding a wider range of organizational roles (Conger & Pearce, 2003; Day et al., 2006; Gronn, 2002; Huusko, 2007; Locke, 2003; Podsakoff, MacKenzie, & Fetter, 1993). This distributed perspective has received particular emphasis in the literature on school leadership due to characteristics of schools as "team-based" (Day et al., 2006; Gronn, 2002, 2009), "loosely-coupled" organizations (March, 1978; Ogawa & Bossert, 1995; Weick, 1976). In combination with recent changes originating in the institutional environment (e.g., increasing accountability, work intensification, role differentiation), this has led to a growing *prescriptive* as well as *descriptive* interest in, the distribution of leadership in schools (Gronn, 2002, 2003; Grubb & Flessa, 2009; Leithwood et al., 2009; Spillane, 2006).

Scholars suggest that the impact of distributed leadership in schools is achieved through improved communication of mission and goals, better alignment of resources and structures to support students, more active engaged professional learning among staff, and the ability to maintain a focus on innovations in teaching and learning by those responsible for implementation (Fullan, 2006; Gronn, 2002; Harris, 2003; Leithwood et al., 2006, 2009; Robinson, 2007; Timperly, 2009). In this study, distributed school leadership was conceived to include a sustained focus on school improvement, support for shared governance, involvement in resource allocation, and to encompass participation from the principal, grade-level heads, teachers, and community representatives on the school's leadership council. We focus on the effects primarily, rather than on the formal/informal leadership arrangements.

1.3.3. School improvement capacity

A substantial body of research has found that leadership effects in schools are mediated by the school's academic and social organization (Hallinger & Heck, 1996a; Leithwood et al., 2004). For the purposes of this study, we refer to this mediating factor as the school's capacity for improvement. This factor is defined from a set of discrete variables that have emerged from several decades of research on school effectiveness and improvement (Teddlie & Reynolds, 2000). The specific observed indicators comprising this factor include the quality of 1) the school's implementation of the state's curricular standards, 2) academic expectations for students, 3) sustained focus on academic improvement, 4) resource support that enables action, 5) continuous professional learning, 6) open communication, and 7) parent support for student learning. We view this set of observed indicators measuring school improvement capacity as *reflective* indicators (versus *formative* indicators) of an underlying process (e.g., see Jarvis, Mackenzie, & Podsakoff, 2003).[1] Tapping into this underlying process with multiple measures and time points provides a valuable way to monitor evolving organizational work structures proximal to student learning across a large number of cases.

In this model, leadership is proposed to achieve its effects on academic outcomes indirectly through building the school's professional capacity and by maintaining a focus on improvements in teaching and learning. This model assumes that changes in distributed leadership and capacity for improvement manifest themselves in latent changes to teachers' practices and students' experiences. We represent this relationship in the *within-schools* portion of the model in Fig. 1 with a dotted oval and arrows, since we do not have direct classroom measures in this study. We illustrate the proposed impact of these school level, latent relationships on classrooms and students with a broad arrow extending from the school level to the classroom and (by association) individual student level of the data hierarchy. Even though we do not measure *classroom changes* directly in this analysis, we assume that changes to teacher classroom behaviors will be responsible for changes we may observe in student growth rates (Cohen & Hill, 2000; Creemers, 1994; Hill & Rowe, 1996; Lee & Bryk, 1989). Because we utilized information from teachers and triangulated it with similar information from students and parents, we believe the data provide a reasonable means to test the proposed conceptual model.

2. Research focus and hypotheses

As noted, this research is grounded in three related problems identified in the literature. First, there have been few large-scale empirical studies of how leadership contributes to school improvement. Second, researchers have yet to provide an empirical test of the proposition that the small indirect leadership effects on student learning found in cross-sectional studies may be larger when examining changes in school improvement capacity and student learning over time. Third, influential scholars in the UK (e.g., Day et al., 2006; Gronn, 2009; Harris, 2003), North America (Fullan, 2006; Leithwood et al., 2009; Spillane, 2006), and Austral-Asia (Mulford, 2007; Robinson, 2007; Timperly, 2009) have identified the need for research that examines the impact of distributed leadership on the school organization as well as student learning.

The broad goal of the study was to explore the contributions of distributed leadership to school improvement capacity and growth in student learning. We advance three hypotheses for testing the model proposed in the paper. Our hypotheses reflect our interest in exploring the effects of school leadership within a mediated-effects model (see Fig. 1).

[1] One empirical test of the validity of this conceptualization is whether the construct is affected by the removal or addition of its observed indicators. In contrast, in a formative model, the indicators are viewed as causing the composite; that is, changes in the indicators result in a change in the composite under investigation. Composite measures are not corrected for measurement error. In this type of measurement model, replacing an indicator can have a considerable effect on the composition of the composite. We conducted several preliminary empirical tests to verify our measurement model contained reflective as opposed to formative indicators underlying the measurement of the initial-status and change constructs (Coltman et al., 2008; Diamantopoulos, 2005). First, reflective indicators should be highly correlated. For the school capacity factor, correlations between the seven subscales ranged from 0.6–0.9. For the distributed leadership construct, we compared the perceptions of students, parents, and teachers. Correlations ranged from 0.3 to 0.4. Second, in a reflective measurement model, the explanatory power of the constructs can be tested by removing items. We tested this by removing subsets of items (3 at a time) from the school capacity factor. The resulting estimates of the latent capacity level and shape factors on achievement and growth, respectively, were consistent with the output reported in Table 4. Similarly, the relationship between distributed leadership and school capacity was unchanged. Finally, we repeated our analyses using student perceptions and then parent perceptions in place of teachers' perceptions of school capacity building and distributed leadership. Results were almost identical.

The first hypothesis (H_1) proposes that teacher perceptions of initial school improvement capacity will be positively related to initial levels of student achievement. Based upon prior research, we suggest that the school improvement capacity factor represents an "alterable variable" that leaders can shape to improve student performance outcomes.

Our second hypothesis (H_2) proposes that changes in school improvement capacity over time will result in measurable changes in students' growth rates in reading and math. This hypothesis tests the dynamic portion of the model in which we assess patterns of change in school improvement capacity and subsequent growth in student performance over time.

Our third hypothesis (H_3) proposes that leadership effects on student learning outcomes will be indirect, operating through the school improvement capacity construct, rather than direct (shown in parentheses in Fig. 1). We propose that these indirect leadership effects on learning outcomes will account for significant differences between schools in their initial achievement levels as well as their subsequent rates of growth.

3. Method

This study employed a non-experimental, post-hoc, longitudinal design (Campbell & Stanley, 1966; Kerlinger, 1986). Although superior to cross-sectional designs for this type of research, longitudinal studies cannot fully resolve the direction of causality between variables, (Cook, 2002). The major threat to validity in longitudinal, non-experimental research comes from uncontrolled or confounding variables.

To test the proposed model and associated hypotheses, survey data were collected from students, parents, and teachers in elementary schools over a four-year period of time. The survey was administered to all certificated staff, all grade five students, and a random sample of parents (i.e., approximately 20% across grade levels). Because teachers are well positioned to understand the school's curriculum, instructional expectations and routines, and are in contact with students and parents regularly, we decided to capture potential changes in leadership and academic processes using the surveys given to each school's teachers on three occasions. However, we also re-ran our analyses with the parent and student data to extend our model's generalizability. Teacher return rates for the three periods of data collection in this study were 73.4%, 76.4%, and 75.6%, respectively.

When surveys are repeated over time with a high level of consistency between items, sequential measures may be used to estimate changes that occur in a population (Davies, 1994). Data from teachers were collected in years one, three, and four. Data on individual student achievement were collected in years two, three and four. We note in passing that unequal spacing of observations and nonlinearity can be incorporated into a LCA model without compromising the quality of data analysis (Raykov & Marcoulides, 2006).

3.1. Sample

A sample of 197 elementary schools was randomly selected from the population of elementary schools in a western state in the United States. A longitudinal cohort of 13,391 third-grade students within the schools (Mean = 92.18, SD = 43.57) participated in the study. Student demographics included in the model are summarized in Table 1. Student socioeconomic status (SES) was estimated by determining participation in the state's federally-funded lunch program. Forty-five percent of the students in the study participated, which is consistent with the reported 42% of public school students in the state who qualified for free or reduced-cost lunch (National Center for Educational Statistics, 2000). Fourteen percent of the students entered the school system after the first year of the study, and 16% changed schools. One advantage of the LCA approach is that missing data and student mobility can be incorporated directly into the analysis, which reduces parameter bias that can result from eliminating these students (Peugh & Enders, 2004).[2]

3.2. Operational definition and measurement of variables

Here we describe how the main conceptual variables defined earlier were operationalized, as well as the measurement properties of the scales. We note that the research relied on secondary data collected by the state's Department of Education. Measurement of variables used to define leadership and school improvement capacity, therefore, was subject to limitations that would not have been present had we developed the measures ourselves.

3.2.1. School context indicators

In addition to the student background variables, context indicators describe initial school conditions during the first year of the study (2002–03), unless otherwise noted. *School size* was defined as the number of students enrolled for the school year. *Student composition* was defined as a composite variable by combining several relevant student background variables to create a weighted school indicator (using principal components analysis). The variables included the percentage of children receiving free or reduced lunch, percentage of students receiving English language (ELL) services, and the percentage of students defined by the state as *underrepresented* in higher education by race/ethnicity. We included this latter indicator because these groups of students

[2] Maximum likelihood estimation is based on available data points, and subjects do not need to have complete data. Partial data actually contribute to the estimation of the model's parameters by implying probable values for missing scores via the correlations among variables. Expectation maximization, a common method for obtaining ML estimates with incomplete data, treats the model parameters (rather than the data points themselves) as missing values to be estimated and borrows information from the existing data at successive iterations until differences between covariance matrices generated are trivial.

Table 1
Descriptive statistics for within-school variables in the model (N = 13,391).

Variable name	*M*	*SD*	Minimum	Maximum
Student outcomes				
Read 2004 (ICC = 12.5%) [a]	245.927	65.523	100.00	500.00
Read 2005 (ICC = 14.7%)	273.127	68.723	100.00	500.00
Read 2006 (ICC = 12.6%)	281.542	73.028	100.00	500.00
Math 2004 (ICC = 14.8%)	217.115	63.432	100.00	500.00
Math 2005 (ICC = 15.1%)	234.998	68.424	100.00	500.00
Math 2006 (ICC = 15.3%)	250.829	66.324	100.00	500.00
Student background				
Low SES	0.45	na	0.00	1.00
English services	0.07	na	0.00	1.00
Special education	0.11	na	0.00	1.00
Female	0.49	na	0.00	1.00
Underrepresented	0.50	na	0.00	1.00
Changed schools	0.16	na	0.00	1.00
Entered year 2/3	0.14	na	0.00	1.00

[a] Intraclass correlation (ICC) refers to the variance in outcomes between schools.

are more heavily concentrated within certain schools in the state. Larger positive values represent school settings where these percentages of students were higher. *Teaching staff stability* was defined as the percentage of teachers who had been at the school for five years (i.e., assessed in year 4). *Principal stability* was defined as whether the same principal (coded 1, else = 0) was at the school during the four years of the study.

3.2.2. Distributed leadership and school improvement capacity

The distributed leadership and school improvement capacity constructs were defined from the survey items. The items were measured on five-point, Likert-type scales. Indicators were expressed as the percentage of positive agreement with each statement. Higher percentages reflect more favorable perceptions about the school's learning environment. Cronbach's alpha (α), a measure of internal consistency, was used to assess the reliability of each subscale.

Distributed leadership was measured by a subscale describing teacher perceptions of leadership from a variety of sources within the school ($\alpha = 0.82$). The stem used for this item was “To what extent does school leadership...” The survey instructions explicitly noted that leadership was not limited to the principal.

The survey items were designed to reflect three specific aspects of distributed leadership: *school improvement* (i.e., To what extent does school leadership: Make decisions to facilitate actions that focus the energies of the school on student achievement and school-wide learner outcomes; Empower staff and students; Encourage commitment, participation and shared accountability for student learning?); *school governance* (i.e., …Adopt governance guidelines which are consistent with the school's purpose and support the achievement of the state standards and the school-wide learner outcomes?); and *resource management and development* (i.e., …Allocate available resources in a manner that sustains the school program and are used to carry out the school's purpose; use assessment results as the basis for the allocation and use of resources?). Factor scores describing the measurement of the leadership factor on each of the three occasions (summarized in the results section) were saved and used to define the LCA model of distributed leadership.

The factor that we named school improvement capacity ($\alpha = 0.95$) was formed by combining seven subscales. Preliminary data analysis (summarized in the results section) treated these subscales as discrete variables in order to examine their psychometric properties in describing the capacity of schools to improve over time. Factor scores were subsequently saved and used to define the LCA model of school capacity improvement at each of the three occasions. The subscale alphas and items comprising each of the subscales were as follows.

- *Standards emphasis and implementation* ($\alpha = 0.91$; 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);
- *Focused and sustained action on improvement* ($\alpha = 0.83$; 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 school-wide and I am involved in the school improvement process);

- *Quality of student support* ($\alpha = 0.85$; 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 offered to students);
- *Professional capacity of the school* ($\alpha = 0.80$; 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);
- *School communication* ($\alpha = 0.88$; School employs a wide range of strategies to ensure parent involvement; open communication among staff; open communication exists between school staff and parents; school responds to parent concerns; school keeps parents informed; I encourage and welcome parents to come to my classroom);
- *Stakeholder involvement* ($\alpha = 0.80$; Parents participate in important decisions about their children's education; school involves parents in classrooms such as tutoring students or checking homework; school encourages parent involvement in a variety of ways); and
- *Student safety and well being* ($\alpha = 0.82$; The school is orderly and supports learning, school staff shows that they respect and care about students, students can receive extra help and support when needed).

3.2.3. Reliability and validity of the instrument

Various forms of the survey instrument employed in this study have been used in previous research studies. This research has shown the subscales defining the school improvement capacity factor to be reliable and valid. The internal consistency estimates obtained meet common standards for survey research (i.e., with α coefficients of 0.80 or above).

Validity of the scales has been assessed for both face and predictive validity. For example, prior studies found a significant relationship between the quality of schools' improvement capacity measured by the survey and sixth-grade student achievement in reading and math (Heck, 2000, 2006), Standardized effects for explaining student achievement levels ranged from 0.08 to 0.34 across five different years of achievement data and from 0.22 to 0.31 on student growth rates across multiple student cohorts. These estimates are similar to the 95% confidence intervals reported for estimates obtained in this study (see Table 4). Prior studies therefore provide consistent evidence of the instrument's reliability and validity.

3.2.4. Math and reading outcomes

Math and reading scores for the cohort of students in the study were collected over three successive years (third–fifth grades). The reading and math tests were constructed in relation to state-developed curricular goals. The tests consist of constructed-response items and standardized test items from the Stanford Achievement Test (Edition 9).

For reading, there were three curricular strands consisting of 47 items (i.e., comprehension process, conventions and skills, and literary response and analysis). For math, there were five strands consisting of 52 items (i.e., number and operation; measurement; geometry and spatial sense; patterns, functions and algebra; data analysis, statistics, and probability). Resulting test scores (re-scaled from 100 to 500) consider patterns of right, wrong, and omitted responses, and item difficulty over successive years. The scores were equated across years to enable the measurement of academic growth.

3.3. Data analysis

Analysis of the data proceeded in several steps. First, we examined changes in the school improvement capacity and distributed leadership factor scores over time. We used the multiple-group capacity of SEM to test the fit of the subscales to the factors across the three measurement occasions (Raykov & Marcoulides, 2006). This analysis was conducted to establish the consistency (i.e., reliability) and validity of our conceptualization of distributed leadership and school improvement capacity over several measurement occasions. More specifically, we wanted to determine whether, in fact, the constructs were measured consistently over time and the extent to which schools improved in their capacity to provide distributed leadership and quality educational practices over the four-year period under study.

Second, we investigated our proposed multilevel latent change model. In the SEM approach to examining individual and organizational change, repeated observations on individuals over time (yt) can be expressed as a type of confirmatory factor analysis (or measurement model), where the level (intercept) and shape (growth) of latent factors are measured by the multiple indicators of y. Our proposed model involves monitoring changes in student academic outcomes at two levels (i.e., within and between schools) and changes in leadership and improvement capacity at the school level. The period of time examined in this study was approximately four academic years. Student-level variables in the model were grand-mean centered, which results in school-level estimates of achievement and growth that have been adjusted for differences in student background within schools. School-level variables (except principal stability) were also centered on their respective grand means for the sample of schools.

We chose to define the distributed leadership and school capacity level factors as *initial status* factors in order to develop a year-1 baseline from which to measure subsequent changes in school leadership and academic processes (i.e., measured again during year 3 and year 4). The initial-status student achievement factor was measured in year 2 and, subsequently, achievement was also measured in year 3 and year 4 (when the cohort was in grades 4 and 5). Finally, we tested the efficacy of our proposed theoretical model highlighting the indirect leadership effects on initial school outcomes and subsequent improvement in learning (through school capacity for improvement) against a more general model that proposed both direct and indirect leadership effects on outcomes by examining the change in chi-square ($\Delta\chi^2$) between models. We also use $\Delta\chi^2$ to test the equality of the size of the leadership and improvement capacity effects in accounting for initial achievement levels and growth over time.

We provide further technical details on the specification and testing of the model in the footnotes.[3] Here we do wish to highlight the special capability of our LCA model to incorporate missing data, unequal spacing of successive measurement occasions, multiple trajectories, and nonlinear effects associated with accelerating or decelerating change. This makes LCA quite flexible for examining different types of organizational change, including parallel changes (where several processes are changing simultaneously) as well as situations where the pattern of change over time is less uniform (e.g., decline followed by rise or vice versa, nonlinear change).

4. Results

The data analysis procedures included tests of the conceptual model as well as the three hypotheses. Descriptive statistics for the variables are provided in Tables 1 and 2. *Within schools*, Table 1 summarizes change in students' reading and math scores between third and fifth grades (i.e., approximately 35.6 points in reading and 33.7 points in math). Intraclass correlations, which describe variance in student achievement attributable to differences between schools, ranged from 12.5% to 15.3% in reading and math, consistent with previous multilevel research on school effects (Hill & Rowe, 1996). Table 2 indicates average yearly growth rates were 18.77 scale score points in reading (SD = 17.4) and 16.52 (SD = 15.5) points in math.

In Table 3, the standard deviations associated with the observed indicators and minimum and maximum percentages suggest considerable variability *between schools* in teachers' perceptions regarding the variables used to define initial levels of distributed leadership and school improvement capacity. Also important to our subsequent analyses, we note that almost one-third (0.31) of the schools had the same principal over the four-year period.

Tests of our proposed model were conducted with Mplus 5.1 (Muthén & Muthén, 1998–2006). We first determined whether schools actually changed in leadership and improvement capacity over time. To conduct this initial analysis of measurement invariance, we used the multiple-group capacity of Mplus to examine the fit of the subscales to the factors across the three measurement occasions (Raykov & Marcoulides, 2006).

Adequacy of the consistency in measuring these processes simultaneously over time is determined by the model fit indices. The standardized root mean square residual (SRMR) describes the average magnitude of model residuals. Values near 0.05 or lower generally indicate an adequate fit of the model to the data (Marcoulides & Hershberger, 1997). The Comparative Fit Index (CFI) compares the fit of the proposed model against a type of baseline (non-fitting) model, with values near 0.95 providing evidence of an adequate model fit (Marcoulides & Hershberger, 1997). In this initial analysis, the SRMR was 0.071 and the CFI was 0.972.

The factor loadings and alpha coefficients for the school capacity factor on each measurement occasion are summarized in Table 3. To examine whether perceptions changed over time, the successive factor means can be simultaneously tested (i.e., with

[3] Matrices and vectors facilitate the specification of models that can be relatively complex (e.g., containing both observed and latent variables and relationships between variables that indicate both direct and indirect effects). The latent change model to represent individual i at time t can be written as

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

where y_{it} is a vector of outcomes for individual i at time t $(y_{i1}, y_{i2}, ..., y_{iT})'$, v_t is a vector of measurement intercepts, Λ_t is a p x m design matrix representing the change process, η_i is an n-dimensional vector of latent variables, $(\eta_{0i}, \eta_{1i} ..., \eta_{pi})'$, K is a $p \times q$ parameter matrix of regression slopes relating x_i covariates $(x_{1i}, x_{2i} ..., x_{pi})'$to the latent factors, and ε_{it} represents time-specific errors which are contained in the theta covariance matrix (Θ). The factor loadings for the latent factors (i.e., two level and two shape factors) are defined in the Λ_t factor loading matrix. For achievement fixing the loadings of the three measurement occasions on the level factors to 1.0 ensures that they are interpreted as *true* (error free) estimates of students' math and reading achievement levels. In LCA, the possibility that students' growth trajectories are nonlinear can be incorporated in the model through the coding scheme for the shape factor (i.e., 0, 1, *). The asterisk indicates a free parameter which can then be estimated in fitting the model. The interval 0 to 1 represents a linear portion in the model describing the change between year 2 and year 3 in the study. The last growth interval represents any nonlinear change that might be present. This coding strategy is also appropriate for handling the unequal spacing of measurement occasions (i.e., as for our measures of distributed leadership and school improvement capacity). Coding the first interval as 0.0 ensures that the level factor represents students' achievement levels at the beginning of the underlying developmental process under investigation (see Raykov & Marcoulides, 2006, for further discussion). The leadership and school improvement capacity trajectories were defined in the same manner (0, 1,*) to incorporate possible nonlinear change in the latter portion of the study. The structural part of the SEM analysis can then be used to investigate the effects of covariates or other latent variables on the latent change factors. For example, for math, we can model variability in level (η_{0i}) and shape (η_{1i}) latent variables as a function of one or more covariates (x_i) plus error:

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

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

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 (equations identical to Eqs. (2) and (3) would be repeated for the reading portion of the model). Once the overall latent curve model has been defined through relating the observed variables to the latent factors that represent the change process (as in Eqs. (1)–(3)), it can be further divided into its respective individual-level measurement and structural models and its organizational-level measurement and structural models. The within- and between-group measurement occasions will load on their corresponding latent level and shape factors through their between-group (Λ_{tB}) and within-group (Λ_{tW}) factor loading matrices. Similarly, factor variances and covariances (Ψ), structural parameters (B), and residual variances (Θ) will have their corresponding within- and between-group matrices. A vector of intercepts (α) is also defined for the between-groups model (and within-group intercepts in Eq. (1) are set to 0). In contrast to achievement and growth, the leadership and school capacity change processes are defined only at the school level.

Table 2
Descriptive Statistics for Between-School Variables in the Model (N = 197).

Variable name	*M*	*SD*	Minimum	Maximum
Context				
Enrollment	468.98	212.99	62.00	1278.00
ELL (%)	8.45	9.02	0.00	61.00
Low SES (%)	50.49	22.63	0.00	97.00
Underrepresented mean (%)	51.16	23.97	3.00	97.00
School composition	−0.03	0.95	−1.98	2.31
Staffing				
Same principal	0.31	na	0.00	1.00
Staff stability (%)	57.28	14.13	9.52	93.33
School achievement				
Initial read Level	217.88	34.01	127.24	324.45
Initial math level	247.45	35.86	108.33	294.09
Read growth rate	18.77	17.44	−25.15	92.87
Math growth rate	16.52	15.51	−11.91	76.52
Initial distributed leadership				
Leadership (%)	75.01	11.52	33.03	98.59
Leadership factor scores				
Year 1	0.00	0.12	−0.43	0.23
Year 3	0.03	0.14	−0.45	0.26
Year 4	0.02	0.14	−0.44	0.25
Initial school improvement capacity (%)				
Learning standards (%)	87.10	6.32	69.14	98.67
Student support (%)	78.48	10.93	37.63	98.77
Capacity (%)	74.53	11.82	40.01	99.11
Communication (%)	81.16	10.78	32.56	98.48
Focused school Improvement (%)	78.41	11.26	47.22	97.35
Involvement (%)	83.31	9.89	39.06	98.24
Safety/well being (%)	84.91	10.26	46.51	98.91
Improvement capacity factor scores				
Year 1	0.00	0.24	−0.70	0.24
Year 3	0.07	0.23	−0.40	0.26
Year 4	0.09	0.24	−0.47	0.28

t-tests) against the initial factor mean ($\overline{X}_1 = 0.00$, SD = 1), which has the advantage of equating the multiple sets of scores to a common metric. The results suggested that on average schools increased their improvement capacity over time (i.e., $\overline{X}_2 = 0.07$; $\overline{X}_3 = 0.09$). Although the factor score metric does not reveal the magnitude of the change, the difference between time 1 and time 2 was statistically significant ($t = 4.83$, $p < .01$). We also examined changes in the distributed leadership factor (which is comprised of one observed scale). The estimated means suggested leadership perceptions increased significantly initially ($t = 2.34$, $p < .05$), then dipped slightly during the last interval.

4.1. Testing the proposed model of leadership and improvement

Our initial analysis established that schools, on average, changed in their distributed leadership and school improvement capacity as perceived by their teachers. We next tested our proposed conceptual model against the data. Estimates of model fit

Table 3
Factor loadings on school capacity factor and distributed leadership factor and tests of significance of the difference in factor means over time.

Variable name	Time 1	Time 2	Time 3	t-test
School capacity				
Learning standards	0.962	0.957	0.801	
Student support	0.978	0.977	0.850	
Capacity	0.948	0.949	0.804	
Communication	0.990	0.991	0.919	
Focused school Improvement	0.997	0.977	0.912	
Involvement	0.976	0.975	0.663	
Safety/well being	0.884	0.896	0.665	
Cronbach's alpha	0.930	0.940	0.950	
Standardized factor means	0.000	0.073 [a]	0.092 [a]	4.831
Distributed leadership				
Leadership subscale	1.000	1.000	1.000	
Standardized factor means	0.000	0.029 [a]	0.020 [a]	2.335

[a] Factor mean is significantly different from initial mean.

Table 4
Standardized estimates, 95% confidence intervals (CI), and power of variables explaining student achievement and growth.

	Standardized coefficients				
	Read	CI	Math	CI	Power
Model for achievement					
School context					
Enrollment	0.09	(−.05, .22)	0.08	(−.06, .22)	0.20
Student composition	−0.04	(−.21, .14)	−0.02	(−.20,−.15)	0.27
Student background					
Female	0.05 [a]	(.03, .07)	−0.01	(−.03, .01)	0.73
Low socioeconomic status (SES)	−0.04 [a]	(−.06,−.02)	−0.04 [a]	(−.06,−.02)	1.00
English language learners (ELL)	−0.24 [a]	(−.26,−.22)	−0.24 [a]	(−.26,−.21)	1.00
Special education	−0.20 [a]	(−.22,−.18)	−0.21 [a]	(−.23,−.19)	1.00
Underrepresented	0.03 [a]	(.00, .06)	0.04 [a]	(.01, .06)	1.00
Hypothesis 1					
Improvement capacity	0.13 [a]	(.02, .73)	0.15 [a]	(.05, .56)	0.84
Hypothesis 3					
Leadership (indirect effect)	0.02 [a]	(.00, .04)	0.02 [a]	(.00, .04)	0.41
Model for growth					
School context					
Enrollment	−0.15 [a]	(−.30, .00)	−0.14	(−.30, .01)	0.72
Student composition	−0.15	(−.33, .02)	−0.16 [a]	(−.32, .00)	0.74
Staff stability	−0.04	(−.13, .05)	−0.01	(−.09, .12)	0.12
Student background					
Female	0.02 [a]	(−.00, .04)	0.02 [a]	(.00, .04)	0.29
Low SES	−0.05 [a]	(−.07,−.03)	−0.05 [a]	(−.07,−.03)	1.00
ELL	0.10 [a]	(.08, .12)	0.13 [a]	(.11, .15)	1.00
SPED	−0.07 [a]	(−.10,−.05)	−0.06 [a]	(−.08,−.04)	1.00
Underrepresented	−0.11 [a]	(−.14,−.09)	−0.11 [a]	(−.14,−.08)	1.00
Changed schools	−0.08 [a]	(−.10,−.06)	−0.08 [a]	(−.10,−.06)	1.00
Hypothesis 2					
Change in capacity	0.20 [a]	(.08, .32)	0.26 [a]	(.13, .39)	0.86
Hypothesis 3					
Change in leadership (indirect effect)	0.10 [a]	(.03, .16)	0.13 [a]	(.05, .20)	0.71

[a] $p<.05$.

(CFI = 0.95; SRMR within schools = .02; SRMR between schools = .06) indicated that the proposed model provided a plausible representation of the data (see Table 4 and Fig. 2).

Table 4 summarizes the results concerning variables that explained differences in average initial school achievement levels (i.e., when students were in third grade) and annual average growth rates.[4] Fig. 2 provides further information about school-level mediating relationships in the proposed model. The coefficients are standardized; this indicates the relative size of each variable's effect[5] with the significance level set at $p = 0.05$. As Hedges (2008) notes, when reporting effect sizes, it is desirable to include estimates of uncertainties (e.g., standard error or confidence intervals). We provide confidence intervals and an estimate of power with respect to each parameter in Table 4. It is also important to note that when interpreting effect sizes the level of analysis matters in multilevel populations. More specifically, 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 (Hedges, 2008). Therefore, it is best to consider specific effects in relation to other observed (or known) effects at each level of the data hierarchy.

Within schools, in Table 4 all of the student background variables were statistically significant in explaining students' *initial achievement levels* in reading, and all but gender were statistically significant in explaining initial achievement levels in math ($p<.05$). Similarly, all of the background variables were significantly related to *growth in reading and math*. Consistent with prior research, therefore, the results suggests differences in student learning levels and growth associated with their background characteristics (Lee & Bryk, 1989).

Between schools, however, neither student enrollment nor student social composition was related to initial achievement levels in reading and math. Social composition was significantly and negatively related to student growth rates in math (standardized $\gamma = -0.16$, $p<.05$). This result can be interpreted as students in schools 1-SD below the grand mean in social composition would have about 0.16 SD larger growth rates in math compared with the growth rates of students in schools on the grand mean for social

[4] R-square coefficients for the within-school model were 0.32 for math level, 0.38 for reading level, 0.08 for math growth, and 0.07 for reading growth. For the between-schools model, the coefficients were 0.54 for initial math level and 0.73 for initial reading level (with leadership and process variables accounting for about 5% of the initial achievement variability). About 15% of the total variation in academic growth lies between schools. The R-square coefficients were 0.21 for math (with context indicators accounting for about 10% of this variability and process variables 11%) and 0.13 for reading (with context indicators accounting for 3% of this variability and process variables about 10%). These R-square estimates are without initial achievement status, which alone accounts for about 60% of the between-school variation in growth for each outcome.

[5] In Mplus, variables are standardized by the variances for latent and observed variables within each level of the data hierarchy, which is useful in determining how much variance is explained at each level.

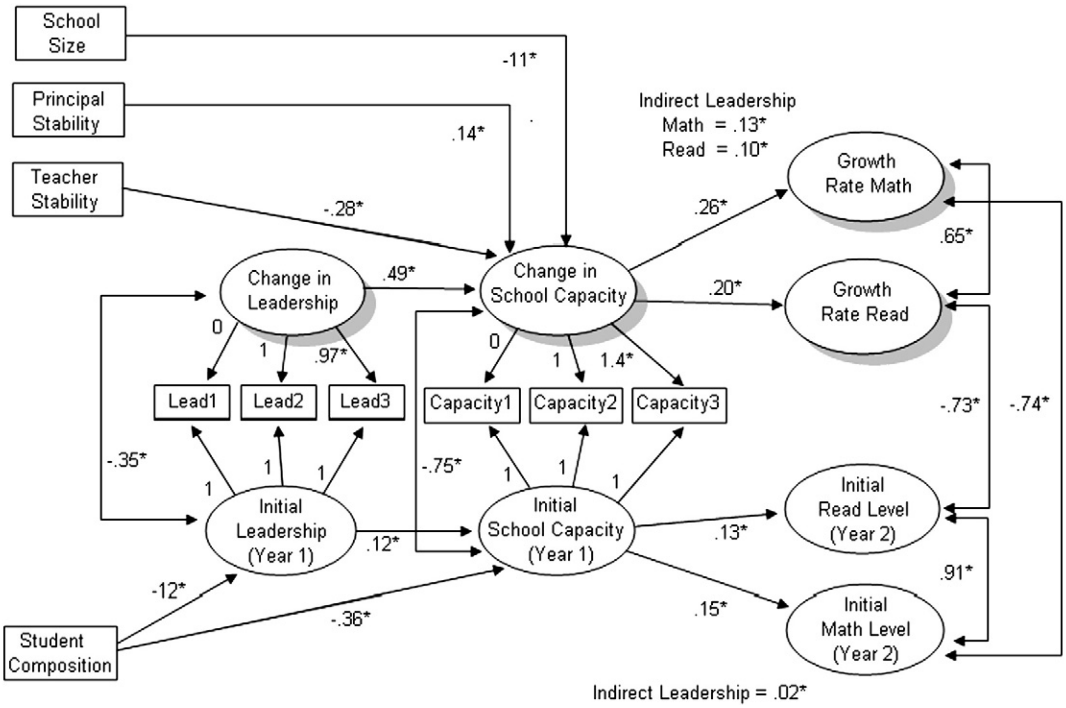

Fig. 2. Between-school standardized effects (*$p<.05$).

composition. School size was significantly and negatively related to student growth rates in reading (standardized $\gamma = -0.15$, $p<.05$).

Fig. 2 indicates several significant effects of context variables on distributed leadership and school improvement capacity, which are not summarized in Table 4. For example, student composition was directly related to initial school improvement capacity (standardized $\gamma = -0.36$, $p<.05$) and perceptions of initial distributed leadership (standardized $\gamma = -0.12$, $p<.05$). These results suggest teacher perceptions of school improvement capacity and distributed leadership were more positive in schools 1-SD below the grand mean in social composition (e.g., percentage of students on free or reduced lunch) compared with teacher perceptions in schools at the grand mean for social composition. This latter finding was consistent with scholars' contention that school context can influence leadership actions (Hallinger & Murphy, 1986; Leithwood et al., 2004, 2006; Teddlie et al., 2000).

Teacher perceptions about improvement capacity were more positive in smaller schools (standardized $\gamma = -0.11$, $p<.05$) compared with teacher perceptions in schools at the grand mean for school size. Perceptions about improved school capacity were inversely related to the stability of teaching staff. Teachers in schools that were relatively more stable in terms of teacher turnover perceived less change in the school's capacity for improvement over the four-year period, compared with teachers in schools that had higher turnover. In contrast, in schools where the same principal was present over the period of the study, teachers were more positive about changes in the school's capacity to improve than teachers in schools where there had been principal turnover (standardized $\gamma = 0.14$, $p<.05$).

4.2. Results of hypothesis testing

Our first hypothesis proposed that initial teacher perceptions about school improvement capacity would be positively related to initial student achievement levels. The results in Fig. 2 support this hypothesis for both reading (standardized $\gamma = .13$, $p<.05$) and math (standardized $\gamma = .15$, $p<.05$). Table 4 also provides these estimates along with confidence intervals and power for detecting each separate effect.

Our second hypothesis proposed that changes in teachers' perceptions over time about their school's improvement capacity would be positively related to growth rates in student achievement in reading and math. Fig. 2 indicates that this hypothesis was supported for growth rates in reading (standardized $\gamma = 0.20$, $p<.05$) and math (standardized $\gamma = 0.26$, $p<.05$). In math, this result can be interpreted as a 1-SD increase in change in school capacity above the grand mean for capacity (0.0) would result in a 0.26 SD increase in average student growth rates. We found that these coefficients were significantly larger in size than the coefficients describing the effects of initial capacity on initial achievement levels ($\Delta\chi^2 = 21.46$ and Δ df $= 3$, $p<.05$).[6]

[6] In SEM, hypothesis tests concerning individual parameters can be tested by comparing a model where the parameters to be tested are freely estimated to a second model where they are fixed to the same value. In this case, the $\Delta\chi^2$ was 21.46 (for 3 degrees of freedom), which exceeds the required $\Delta\chi^2$ of 7.82 at $p = .05$.

Moreover, we found that the size of the estimated relationships between school improvement capacity and learning outcomes actually changed over the several measurement occasions. To illustrate this, in Fig. 3, we represent visually the increasing effects, summarized as correlations (r), between improvement capacity and achievement for each measurement occasion ($r = 0.22$ at time 1, represented by the solid line; $r = 0.35$ at time 2, represented by the dotted line; $r = 0.42$ at time 3, represented by the dashed line). The pattern results in increasing regression slopes (and R^2) as shown in the figure. This provides one concrete illustration of nonlinear effects embedded in our change model that are missed in cross-sectional studies.

Our third hypothesis proposed that initial distributed leadership and change in distributed leadership would have significant *indirect* effects on school achievement and growth rates, respectively. This hypothesis actually has two parts. First, it implies that both the initial level and shape factors for leadership and improvement capacity should be related to each other. Our belief was that relationships between initial processes would likely be weaker than subsequent relationships where there has been a state-wide effort to improve the school. Since leadership is often conceptualized as a catalyst for change, it follows that stronger perceptions of leadership would be associated with increased capacity for school improvement.

Second, this hypothesis implies that the combined effects of distributed leadership on the initial achievement and growth outcomes should be indirect rather than direct. In particular, we suggest that indirect effects of leadership are likely to be larger for improvement processes than for initial achievement levels. This follows from the argument advanced that leadership effects on both capacity and learning outcomes will become more noticeable over a period of time.

Regarding the first part of Hypothesis 3, we found that initial distributed leadership was significantly related to initial improvement capacity (standardized $\gamma = 0.12$, $p < .05$). Similarly, a change in distributed leadership was also significantly related to change in school improvement capacity (standardized $\gamma = 0.49$, $p < .05$). The size of the relationship between the change factors was, as expected, significantly larger than the relationship between the level factors describing initial leadership and improvement capacity conditions ($\Delta\chi^2 = 281.19$ and Δ df $= 1$, $p < .05$), a point we shall return to subsequently.

The second part of Hypothesis 3 stated that the indirect effects of distributed leadership on learning outcomes should be significant. As summarized in Table 4 and Fig. 2, we found support for all four of these relationships ($p < .05$). The size of the indirect effect of *initial* distributed leadership on *initial* reading and math achievement levels was very small (standardized $\gamma = 0.02$, $p < .05$). Indeed, given the low power to detect the relationship (0.41), we suggest that these indirect effects might not be replicated in repeated samples taken from the population (Muthén & Muthén, 1998–2006).

Notably, however, the indirect effects of *changes* in distributed leadership on *changes* in student learning were larger (standardized $\gamma = 0.10$ for reading and 0.13 for math, $p < .05$). Moreover, the power for detecting these latter effects on growth, was near the accepted standard 0.80 (i.e., 0.71). This suggests that the relationship would likely be replicated in repeated samples from the population. Although the size of this effect may still appear small, this puts the indirect effects of leadership on student growth in learning about on a par with direct effects of initial school improvement capacity on initial school achievement levels (or school size on reading growth) in our model. These results support the view that the effects of both distributed leadership and improvement capacity increase in terms of their impact when growth (or change) is the performance outcome (accounting for 10–11% of the variation), as opposed to outcome levels at one point in time (accounting for roughly 5% of the variation).

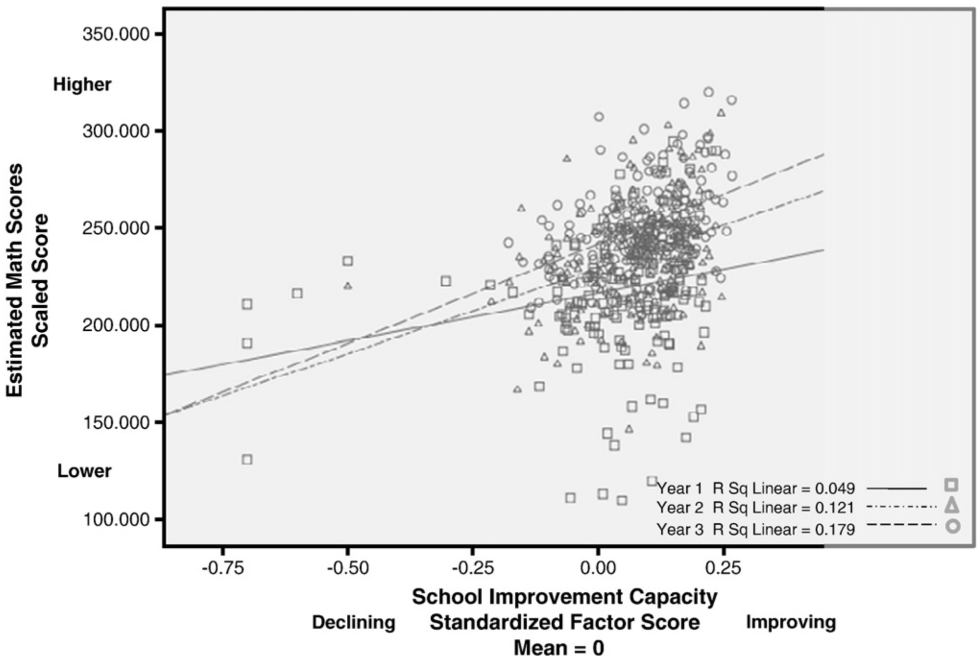

Fig. 3. Increasing linear relationships between school improvement capacity and estimated math scores over time in sample elementary schools.

We can also compare the fit of the proposed model with indirect leadership effects on the academic outcomes against a model having both direct and indirect leadership effects on the outcomes. This comparison model has four more estimated parameters to represent the direct paths between distributed leadership and reading and math initial achievement and between change in distributed leadership and growth rates in reading and math. The fit of the nested models to the data was evaluated by examining the difference in chi-square coefficients between the two models, after appropriate scaling adjustment for non-normality (Muthén & Muthén, 1998–2006). We found the indirect-effects model fit the data better than the model with both direct and indirect effects ($\Delta\chi^2 = 94.55$, $p < .001$) with Δ df = 4. This finding reinforces the theoretical assertion that the effect of leadership on school outcomes is primarily indirect as opposed to direct. Moreover, the finding also supports the conclusion that bivariate studies which explore the direct effects of leadership on student outcomes represent a "dry hole" in this domain of leadership research (Hallinger & Heck, 1996b; Heck & Hallinger, 2005; Robinson, 2007).

5. Discussion

The challenge of improving organizational performance has led to the study of alterable variables that leaders can shape in order to make a difference in performance (Campbell et al., 1993; Kaiser et al., 2008; Steers, 1975). Although effective organizations appear to share similarities (Kotter & Heskett, 1992; Marcoulides & Heck, 1993; Nonaka & Toyama, 2002; Podsakoff et al., 1993), their inherent complexity has made it difficult to establish empirically a causal linkage between changes in leadership or organizational processes and changes in performance over time (Hallinger & Heck, 1996a; Kaiser et al., 2008; Marcoulides & Heck, 1993; Ouston, 1999; Podsakoff, 1994). For example, theorists have suggested that organizations proceed through cycles of innovation and change over time (Huber & Van de Ven, 1995; Kimberly & Miles, 1980). Mapping the contribution of leadership to performance change across organizations located at different points in their own improvement cycles requires sophisticated theoretical models and analytical methods.

The present study proposed to test a longitudinal model of distributed leadership and change aimed at improvements in school performance. We examined how changes in school improvement capacity in key areas were related to changes in student learning. Our thesis was that school leadership is central to a constellation of variables that describe school improvement capacity and which account for differences in student learning between schools. Our model proposed two types of temporal relationships simultaneously in order to capture these effects:

- First, we modeled the initial relationship between distributed leadership, capacity for school improvement, and levels of student achievement at a single point in time.
- Second, we modeled *changes* in these same organizational process variables along with growth in student learning outcomes over several years.

The results lend support for the view that longitudinal models can capture changes in organizational structures and processes that lead to changes in performance (Blalock, 1989; Monge, 1990; Nonaka & Toyama, 2002). LCA modeling enabled the examination of concurrent changes in key organizational variables that were related to a multilevel model of students' growth in reading and math over time. Our analysis incorporated missing data, unequal spacing of measurement occasions, simultaneous change processes at multiple organizational levels, and nonlinear effects associated with accelerating or decelerating change over time. Thus it addresses some common problems that have been associated with investigating organizational change.

First, the results suggested that, on average, perceptions of school improvement capacity increased over time. We note that this research was conducted during a period of aggressive state implementation of educational reforms that sought to foster distributed leadership and support school improvement. Although our research was not designed to assess the effects of the new state policy reforms, the results suggest that the policy may have provided some impetus for schools to change (e.g., increased monitoring of required classroom changes in curriculum).

Second, consistent with our proposed hypothesis (H_1 in Fig. 1), at the beginning of the study existing differences between schools in improvement capacity were statistically associated with differences in initial levels of student achievement in reading and math after controlling for relevant context variables. These findings demonstrate the added value of monitoring the changing professional capacity of schools and its relationship to growth in student learning.

Third, we proposed that changes over time in school improvement processes (e.g., focus on school improvement, academic expectations) would add value in terms of increased rates of student growth in learning (H_2). Proposed relationships for reading and math were statistically significant and more substantial in size than other school effects in the model (e.g., student composition, school size, staff stability). More specifically, changes in key school educational process indicators (i.e., that create supportive conditions for learning) over time accounted for substantial variation (10–11%) in student growth rates between schools.

Previous research on school improvement has not typically addressed the problem of modeling changes in school performance using longitudinal indicators of processes and outcomes. As the initial-status portion of our model suggests, examining processes in organizations at any one point in time provides a very limited snapshot of conditions—as if nothing is "in motion." In those circumstances, it is not possible to represent the separate trajectories that different organizational processes are following over time or to determine the relationships among the variables at other points in time (i.e., before or after the snapshot). This limitation of cross-sectional research may explain why effects attributed to leaders or other educational processes are so varied in size (Hallinger & Heck, 1996a). Moreover, at any point in time it is possible that extraneous variables (e.g., idiosyncratic turnover, community problems, funding changes) may exert an impact on leadership or other variables. Longitudinal data collection and analyses can facilitate the separation of extraneous factors from intended effects.

We monitored changes in schools' distribution of leadership and improvement capacity over time in order to provide a preliminary test of whether upgrading the school's improvement capacity might correspond with increased student learning rates (Fullan, 2006). The findings provide preliminary evidence of the construct validity of measures of change in school improvement capacity and their efficacy for estimating the effects of the schools' efforts to reform their educational practices. This finding is encouraging because it supports the view that intentional efforts to reshape organizational structures and routines can impact change in student performance (Ouston, 1999). Growth in student outcomes, therefore, may be a more valid indicator of school effectiveness than outcomes measured at one point in time (Willms, 1992).

Fourth, we proposed that the effects of changes in distributed leadership would be indirectly related to growth in student learning (H_3). An indirect effect implies that the relationship between leadership and outcomes is mediated by educational practices and strategic changes in those practices. The results were significant for both reading and math. In addition, we noted significant indirect effects between initial levels of distributed leadership and initial student achievement. These initial effects, however, were substantively small and less important. In contrast, indirect effects of distributed leadership on growth rates were substantively larger, thereby suggesting that leadership had a potentially important impact on improvement in these school settings. Model testing further supported a conceptualization of leadership effects as both indirect and unfolding over time, as opposed to being readily observable at any one point in time.

6. Implications

These results provide support for longitudinal studies that examine how changes in patterns of social interaction and structures influence performance growth in organizations (Campbell et al., 1993; Kaiser et al., 2008; Langlois & Robertson, 1993; Podsakoff, 1994; Williams & Podsakoff, 1989). The findings also reinforce the view that organizational processes can be altered through strategic action, be it through leadership or policy intervention (Langlois & Robertson, 1993; Teece, 1982; Yukl, 2006).

With respect to educational organizations more specifically, our results speak to the importance of placing student learning at the center of research on leadership and school improvement (Heck & Hallinger, 2005; Leithwood et al., 2004, 2006; Robinson, 2007). This priority reflects a global concern for fostering improvement in learning results for all students. Bolstered by stronger theory about organizational change, sustained inquiry should address how capacity for school improvement develops and changes over time, and monitor the subsequent impact on growth in student learning. This research should also seek to describe the role of leadership in initiating, facilitating, and sustaining improvement over time (Fullan, 2006; Hall & Hord, 2001; Leithwood et al., 2004, 2006; Ouston, 1999). A critical line of inquiry will involve assessing the nature of leadership and its impact at different points in the improvement cycles of schools (Hallinger, 2003; Leithwood et al., 2006).

Our study is representative of emerging research that seeks to study sources of school leadership beyond hierarchical roles (Day et al., 2006; Gronn, 2002; Harris, 2003; Leithwood et al., 2009; Marks & Printy, 2003; Mulford & Silins, 2003). Scholars and practitioners alike suggest that the changing context of schools requires the development of broader and deeper leadership resources (Barth, 2001; Gronn, 2009; Lambert, 2002; Ogawa & Bossert, 1995). This research is one of the first large-scale studies to explicitly test the practices of distributed leadership against empirical data on school improvement and growth in student learning. We also explored the generalizability of our proposed model using teacher perceptions against student and parent data about school processes and noted considerable consistency.[7]

Leadership effect sizes in this study were consistent with other known school-level variables that have received considerable policy attention (e.g., school size, student composition, staff stability). The evidence therefore suggests that change in distributed leadership can be empirically linked to change in school improvement capacity and subsequent growth in student learning. Importantly, we found the relative size of the effects were significantly larger when change-related variables (e.g., growth in learning outcomes) were the focus of attention than when similar variables were measured at only one time. One reason may be because error is reduced with latent variable modeling of repeated measurement occasions which, in turn, enhances structural relationships between variables (Raykov & Marcoulides, 2006).

The findings also suggest possible directions for studying the effects of leadership development through large-scale longitudinal research. Nations throughout the world have made large investments in school leadership development over the past 15 years (Huber, 2003). Yet, there have been few rigorous evaluations of the impact of those programs either upon leadership practice or the capacity of schools to improve. We suggest that it would be possible to organize a large-scale study along similar lines to this one, but which also incorporates an interrupted time-series design to examine the impact of leadership development on participants and their schools (Campbell & Stanley, 1966; Shadish, Cook, & Campbell, 2002).

The results of this study further suggest that growth in student learning may be a more salient indicator for school accountability than the level of student achievement measured at one point in time. This has potential implications for policy analysts involved in the design of data management and accountability systems used to monitor school performance. At the same

[7] The models were generally consistent, with some variation in the size of the effects observed. For example, for the parent data, we noted the effects of organizational processes on change in achievement were slightly larger than the model with teacher perceptions. The latent process variables defined with teacher and parent data correlated moderately ($r = 0.50$, $p < .05$). We noted that student views about classroom processes were helpful in evaluating the extent to which school improvement activities may have resulted in classroom changes.

time, however, we should also caution that growth trajectories have not yet been studied sufficiently as a dependent variable to evaluate the longer-term trends in school improvement.

Several limitations should be noted when considering the results of this study. First, questions remain about the day-to-day implementation of leadership efforts aimed at improving school improvement capacity. This study did not employ an experimental design and, therefore, can only speculate about causal linkages. Longitudinal, multilevel data collection represents a clear strength of the study; however, additional waves of data would enable a more robust time-series examination of the causal linkages in the proposed change model over time.

Although our research explicitly adopted a multilevel conception, the study did not incorporate observed measures of classroom practice. Thus, the causal link between school improvement capacity and student achievement remains a “black box.” Our model posits that the effects of school improvement capacity are achieved through changes in teacher practice (and teacher perceptions suggested some change took place over time); however, the data did not support a direct test of this assumption. Future research should seek to collect more thorough information about leadership efforts to change teachers' instructional behavior in classrooms (Cohen & Hill, 2000; Wenglinksy, 2002). Moreover, we acknowledge that the type of school-level aggregates employed in this study ignore wide variations in the conditions of learning and teaching that may be very important at the classroom level (Creemers, 1994).

Questions also remain about the temporal sequence underlying associations between leadership, school improvement capacity, and student outcomes. Although the results of the longitudinal analysis reveal preliminary evidence that building capacity for improvement can increase student growth rates, they do not provide complete protection against a selection-bias argument. For example, teachers may perceive improvement capacity more positively in schools that achieve at high levels over long periods of time. Although our study begins to address the importance of temporal relationships in organizational models, further research is needed to refine causal relationships, including possible reciprocal effects between leadership, organizational processes, and outcome variables within longitudinal models.

Finally, caution must be exercised in using SEM applications (e.g., LCA) to test substantive theories. Omitted variables and measurement error are common sources of misspecification that can produce misleading results (Bentler & Bonett, 1980). The psychometric quality of behavioral measurements is generally evaluated in terms of reliability and validity. An important requirement for the evaluation of models using SEM lies in using theoretically appropriate operationalizations of both observed and latent variables. Of course, evidence of validity is often less obvious than evidence for reliability. 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).

The correct use of SEM to define and test models must therefore be considered carefully within the context of several guidelines. These include the theoretical foundation of the model tested, the validity of the measured variables, the nature of the relationships between the observed variables and the latent variables, the direction of causality, and errors in the measurement of latent variables and errors in the overall model (Heck & Thomas, 2009; Raykov & Marcoulides, 2006). For example, our initial examination of the measurement model used to define the latent improvement capacity and leadership factors at three points in time suggested that school improvement capacity and leadership can be reliably measured. We then provided evidence that teachers' perceptions about schools' underlying capacity and distributed leadership changed over time. Further, fit of the proposed theoretical model to the data suggested the information contained in the latent variables was important in determining whether schools were performing at higher or lower academic levels. What remains is to put together various pieces of the puzzle about the landscape of improving schools. Research that continues to focus on student growth as an outcome and recognizes the centrality of distributed leadership and improvement capacity should have potential for furthering our understanding of the role of leadership in facilitating organizational change.

Acknowledgements

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

References

Atwater, L., Dionne, S., Avolio, B., Camobreco, J., & Lau, A. W. (1999). A longitudinal study of the leadership development process: Individual differences predicting leader effectiveness. *Human Relations*, *52*, 1543−1562.

Barth, R. (2001). *Teacher leader* (pp. 443−449). Phi Delta Kappan.

Bass, B. (1985). *Leadership and performance beyond expectations.* New York: Free Press.

Bass, B. (1990). *Bass and Stogdill's handbook of leadership*, 3rd edition. New York: Free Press.

Bass, B., & Avolio, B. (1994). *Improving organizational effectivenesss through transformational leadership.* Newbury Park, CA: Sage.

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.

Bennis, W. (2003). *On becoming a leader.* Boston: Perseus Publishing.

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

Blalock, H. (1989). The real and unrealized contributions of quantitative sociology. *American Sociological Review*, *54*, 447−460.

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

Campbell, J., McCloy, R., Oppler, S., & Sager, C. (1993). A theory of performance. In N. Schmitt, & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35−71). San Francisco: Jossey-Bass.

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

Carroll, T., & Burton, R. (2000). Organizations and complexity: Searching for the edge of chaos. *Computational and Mathematical Organization Theory*, *6*(4), 319−337.

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

Coleman, J., Campbell, E., Hobson, C., McPartland, F., Mood, A., Weinfeld, F., et al. (1966). *Equality of educational opportunity study.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education/National Center for Education Statistics.

Coltman, T., Devinney, T., Midgley, D., & Venaik, S. (2008). *Formative versus reflective measurement models: Two applications of erroneous measurement.* Unpublished paper. Retrieved September 23, 2008 from http://www2.agsm.edu/agsm/web.nsf/AttachmentsByTitle/TD_Formative_Paper/$FILE/Formative+Indicators.pdf

Conger, J., & Pearce, C. (2003). A landscape of opportunities: Future research on shared leadership. In C. L. Pearce, & J. Conger (Eds.), *Shared leadership: Reframing the hows and whys of leadership* (pp. 285−303). Thousand Oaks, CA: Sage Publications.

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

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

Darling Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence.*Education Policy Archives*, *8*(1) Retrieved October 1, 2006 from http://epaa.asu.edu/epaa/v8n1/on

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., & Salasc, S. (2006). Leadership in team-based organizations: On the threshold of a new era. *Leadership Quarterly*, *17*(3), 211−216.

Diamantopoulos, A. (2005). The COARSE procedure for scale development in marketing. *International Journal of Research in Marketing*, *22*(1), 1−9.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(1), 15−24.

Edmondson, A., Roberto, M., & Watkins, M. (2003). A dynamic model of top management team effectiveness. *Leadership Quarterly*, *14*, 297−325.

Foster, R. (2005). Leadership and secondary school improvement: Case studies of tensions and possibilities. *International Journal of Leadership in Education*, *8*(1), 35−52.

Fullan, M. (2006). *Turnaround leadership.* New York: Wiley & Sons.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, *59*, 185−198.

Glover, J., Rainwater, K., Friedman, H., & Jones, G. (2002). Four principles for being adaptive (Part Two). *Organizational Development Journal*, *20*(4), 18−38.

Gresov, C., Haveman, H., & Oliva, T. A. (1993). Organizational design, inertia, and the dynamics of competitive response. *Organization Science*, *4*(2), 181−208.

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

Gronn, P. (2003). *The new work of educational leaders: Changing leadership practice in an era of school reform.* London: Sage Publications.

Gronn, P. (2009). Hybrid leadership. In K. Leithwood, B. Mascall, & T. Strauss (Eds.), *Distributed leadership according to the evidence*. London: Routledge.

Grubb, W. N., & Flessa, J. (2009). "A job too big for one": Multiple principals and other non-traditional approaches to school leadership. In K. Leithwood, B. Mascall, & T. Strauss (Eds.), *Distributed leadership according to the evidence* (pp. 137−164). London: Routledge.

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

Hallinger, P. (2003). School leadership development in global perspective: Future challenges and opportunities. In P. Hallinger (Ed.), *Reshaping the landscape of school leadership development: A global perspective* (pp. 289−300). Lisse, Netherlands: Swets & Zeitlinger.

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. (1996a). The principal's role in school effectiveness: An assessment of methodological progress, 1980–1995. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 723−783). Dordrecht, Netherlands: Kluwer Academic Publishers.

Hallinger, P., & Heck, R. (1996b). 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 and school improvement. In A. Harris, C. Day, D. Hopkins, M. Hadfield, A. Hargreaves, & C. Chapman (Eds.), *Effective leadership for school improvement* (pp. 72−83). London: Routledge/Falmer.

Hawley, W., & Rosenholtz, S. (1984). Good schools: What research says about improving school achievement. *Peabody Journal of Education*, *61*, 117−124.

Hayes, D., Christie, P., Mills, M., & Lingard, B. (2004). Productive leaders and productive learners: Schools as learning organisations. *Journal of Educational Administration*, *42*(4/5), 520−538.

Heck, R. H. (2000). Examining the impact of school quality on school outcomes and improvement: A value-added approach. *Educational Administration Quarterly*, *36*(4), 513−552.

Heck, R. H. (2006). Assessing school achievement progress: Comparing alternative approaches. *Educational Administration Quarterly*, *42*(5), 667−699.

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

Heck, R., & Thomas, S. (2009). *Structural equation modeling*, 2nd edition. London: Routledge.

Hedges, L. (2008). What are effect sizes and why do we need them? U.S. Department of Health and Human Services, Administration for Children & Families, Paper No. 1. Retrieved December 6, 2008 from http://www.acf.hhs.gov/programs/opre/other_resrch/effect_sizes/pres_papers/hedges.html

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

Holmqvist, M. (2003). A dynamic model of intra- and interorganizational learning. *Organization Studies*, *24*, 93−121.

Hooijberg, R., & Schnieder, M. (2001). Behavioral complexity and social intelligence: How executive leaders use stakeholders to form a systems perspective. In S. Zaccaro, & R. Klimoski (Eds.), *The nature of organizational leadership* (pp. 104−131). San Francisco: Jossey-Bass.

Howard, V., McLaughlin, T., & Vacha, E. (1996). Educational capital: A proposed model and its relationship to academic and social behavior of students at risk. *Journal of Behavioral Education*, *6*(2), 135−152.

Huber, S. (2003). School leadership development: Current trends from a global perspective. In P. Hallinger (Ed.), *Reshaping the landscape of educational leadership development: A global perspective* (pp. 273−288). Lisse, Netherlands: Swets and Zeitlinger.

Huber, G., & Van de Ven, A. (1995). Introduction. In G. R. Huber, & A. H. Van de Ven (Eds.), *Longitudinal field research methods: Studying processes of organizational change* (pp. vii−xiv). Thousand Oaks, CA: Sage.

Huusko, L. (2007). Teams as substitutes for leadership. *Team performance Management*, *13*(7/8), 244−258.

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

Jarvis, C., MacKenzie, S., & Podsakoff, P. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(3), 199−218.

Kaiser, R. B., Hogan, R., & Craig, S. B. (2008). Leadership and the fate of organizations. *The American Psychologist*, *63*(2), 96.

Kelly, J., & McGrath, J. (1988). *On time and method.* Newbury Park, CA: Sage.

Kerlinger, F. (1986). *Foundations of behavioral research*, 3rd edition. New York: Holt, Rinehart and Winston, Inc..

Kimberly, J., & Miles, R. (1980). *The organizational life cycle.* San Francisco: Jossey-Bass.

Kotter, J., & Heskett, J. (1992). *Corporate culture and performance.* New York: The Free Press.

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

Langlois, R., & Robertson, P. (1993). Business organisation as a coordination problem: Toward a dynamic theory of the boundaries of the firm. *Business and Economic History*, *22*(1), 31−41.

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

Lee, V., Croninger, R., & Smith, J. (1997). Course-taking, equity, and mathematics learning: Testing the constrained curriculum hypothesis in U.S. secondary schools. *Educational Evaluation and Policy Analysis*, *19*, 99—121.
Leithwood, K., Louis, K. S., Anderson, S., & Wahlsttom, K. (2004). *Review of research: How leadership influences student learning.*: Wallace Foundation Downloaded from http://www.wallacefoundation.org/NR/rdonlyres/E3BCCFA5-A88B-45D3-8E27-B973732283C9/0/ReviewofResearchLearningFromLeadership.pdf on December 19, 2007.
Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership.* Nottingham, England: National College of School Leadership.
Leithwood, K., Mascall, B., & Strauss, T. (2009). New perspectives on an old idea. In K. Leithwood, B. Mascall, & T. Strauss (Eds.), *Distributed leadership according to the evidence* (pp. 1—14). London: Routledge.
Locke, E. (2003). Leadership: Starting at the top. In C. J. Pearce, & C. Conger (Eds.), *Shared leadership: Reframing the hows and whys of leadership* (pp. 271—284). Thousand Oaks, CA: Sage.
Lord, R. (2001). The nature of organizational leadership: Conclusions and implications. In S. Zaccaro, & R. Klimoski (Eds.), *The nature of organizational leadership* (pp. 413—436). San Francisco: Jossey-Bass.
March, J. (1978). The American public school administrator: A short analysis. *School Review*, *86*, 217—250.
Marcoulides, G., & Heck, R. (1993). Organizational culture and performance: Proposing and testing a model. *Organization Science*, *4*(2), 209—225.
Marcoulides, G., & Hershberger, S. (1997). *Multivariate statistical methods: A first course.* New York: Lawrence Erlbaum Associates.
Marks, H., & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370—397.
Marzano, R. J., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results.* Auroroa, CO: ASCD and McREL.
Mathieu, J., Ahearne, M., & Taylor, S. (2007). A longitudinal cross-level model of leader and salesperson influences on sales force technology use and performance. *The Journal of Applied Psychology*, *92*(2), 528—537.
Mohr, L. (1983). The implications of effectiveness theory for managerial practice in the public sector. In K. Cameron, & D. Whetton (Eds.), *Organizational effectiveness: A comparison of multiple models* (pp. 225—239). New York: Academic Press.
Monge, P. (1990). Theoretical and analytical issues in studying organizational processes. *Organization Science*, *1*(4), 406—430.
Mortimore, P. (1993). School effectiveness and the management of effective learning and teaching. *School Effectiveness and School Improvement*, *4*, 290—310.
Mulford, B. (2007). Building social capital in professional learning communities: Importance, challenges and a way forward. In L. Stoll, & K. Seashore Louis (Eds.), *Professional learning communities: Divergence, depth and dilemmas* (pp. 166—180). London: Open University Press and McGraw Hill.
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.
Mumford, M., Zaccaro, S., Johnson, J., Diana, M., Gilbert, J., & Threlfall, K. (2000). Patterns of leader characteristics: Implications for performance and development. *The Leadership Quarterly*, *11*(1), 115—133.
Muthén, L., & Muthén, B. (1998–2006). *Mplus user's guide.* Los Angeles: Authors.
National Center for Educational Statistics (2000). *Monitoring school quality: An indicators report.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
NCLB (2001). *No child left behind act.* Public Law 107-110. Washington D.C.: United States Government.
Nicolaidou, M., & Ainscow, M. (2005). Understanding failing schools: Perspectives from the inside. *School Effectiveness and School Improvement*, *16*(3), 229—248.
Nissen, M., & Levitt, M. (2002). Dynamic models of knowledge-flow dynamics. *Working Paper #76*. Palo Alto, CA: Center for Integrated Facility Engineering, Stanford University.
Nonaka, I., & Toyama, R. (2002). A firm as a dialectical being: Towards a dynamic theory of a firm. *Industrial and Corporate Change*, *11*(5), 995—1009.
Ogawa, R., & Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, *31*(2), 224—243.
O'Toole, J. (1995). *Leading change.* San Francisco: Jossey Bass.
Ouston, J. (1999). School effectiveness and school improvement: Critique of a movement. In T. Bush, L. Bell, R. Bolam, R. Glatter, & P. Ribbins (Eds.), *Educational management: Redefining theory, policy and practice* (pp. 166—177). London: Paul Chapman.
Peugh, J., & Enders, C. (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—19.
Pitner, N. (1988). The study of administrator effects and effectiveness. In N. Boyan (Ed.), *Handbook of research in educational administration*. New York: Longman.
Ployhart, R., Holtz, B., & Bliese, P. (2002). Longitudinal data analysis applications of random coefficient modeling to leadership research. *Leadership Quarterly*, *13*(4), 455—486.
Podsakoff, P. (1994). Quantitative methods in leadership research. *Leadership Quarterly*, *5*, 1—2.
Podsakoff, P., MacKenzie, S., & Fetter, R. (1993). Substitutes for leadership and the management of professionals. *Leadership Quarterly*, *4*, 1—44.
Raudenbusch, S., & Willms, J. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*(4), 307—335.
Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*, 2nd edition. 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 Press.
Robinson, V. (2007). School leadership and student outcomes: Identifying what works and why. *Melbourne: Australian Council for Educational Leaders Monograph No. 41*.
Rutter, M. (1983). School effects on pupil progress. Research findings and policy implications. In L. Shulman, & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 3—41). White Plains, NY: Longman.
Sammons, P., Nuttall, D., Cuttance, P., & Thomas, S. (1995). Continuity of school effects: A longitudinal analysis of primary and seconday school effects on GCSE performance. *School Effectiveness and School Improvement*, *6*, 285—307.
Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.
Sivasubramaniam, N., Murry, L., Avolio, B., & Jung, D. (2002). A longitudinal model of the effects of team leadership and group potency on group performance. *Group & Organization Management*, *27*(1), 66—96.
Sleegers, P., Geijsel, F., & Van den Berg, R. (2002). Conditions fostering educational change. In K. Leithwood, & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 75—102). Dordrecht: Netherlands: Kluwer Academic.
Spillane, J. (2006). *Distributed leadership.* San Francisco: Jossey-Bass.
Steers, R. (1975). Problems in the measurement of organizational effectiveness. *Administrative Science Quarterly*, *20*(4), 546—558.
Stoll, L., & Fink, D. (1996). *Changing our schools: Linking school effectiveness and school improvement.* Buckingham: Open University Press.
Tate, B. (2008). A longitudinal study of the relationships among self-monitoring, authentic leadership, and perceptions of leadership. *Journal of Leadership and Organizational Studies*, *15*(1), 16—29.
Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer.
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.
Teece, D. (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior and Organization*, *3*, 39—63.
Thomas, A. (1988). Does leadership make a difference to organizational performance? *Administrative Science Quarterly*, *33*(22), 401—417.
Timperly, H. (2009). Distributing leadership to improve outcomes for students. In K. Leithwood, B. Mascall, & T. Strauss (Eds.), *Distributed leadership according to the evidence* (pp. 197—222). London: Routledge.
Tushman, M., & Romanelli, E. (1985). Organizational evolution: A metamorphosis model of convergence and reorientation. In L. L. Cummings & B. M. Staw (Eds.), *Research in Organizational Behavior, Vol. 7* (pp. 171—222). Greenwich, CT: JAI Press.

Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, *21*, 1—16.
Wenglinksy, H. (2002). How schools matter. The link between teacher classroom practices and student academic performance.*Education Policy Analysis Archives*, *10*(12) Retrieved from http://epaa.asu.edu/epaa/v10n12/, February, 12, 2007.
Wiley, S. (2001). Contextual effects on student achievement: School leadership and professional community. *Journal of Educational Change*, *2*(1), 1—33.
Williams, L., & Podsakoff, P. (1989). Longitudinal field methods for studying reciprocal relationships in organizational behavior research: Toward improved causal analysis. *Research in Organizational Behavior*, *11*, 247—292.
Willms, J. D. (1992). *Monitoring school performance.* London: Falmer.
Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*, 398—425.
Yukl, G. (2006). *Leadership in organizations.* Paramus: Prentice Hall.