Collaborative Leadership Effects on School Capacity and Student Learning

Paper Presented at the

International Congress for School Effectiveness and School Improvement

Kuala Lumpur, Malaysia, January 2010

Philip Hallinger

Hong Kong Institute of Education

and

Ronald H. Heck

University of Hawaii Manoa

Contact email: hallinger@gmail.com

Researchers have persisted in framing leadership as the driver for change and performance improvement in schools despite convincing theoretical commentary that proposes leadership as a process of reciprocal interaction. Although conceptualizing leadership as a reciprocal process offers leverage for understanding leadership effects on learning, methodological constraints have limited empirical tests of this model. This report focuses on understanding the contribution that changes in collaborative leadership and the school's capacity for educational improvement make on growth in student learning. We compare longitudinal, unidirectional and reciprocal-effects models focusing on change in leadership and learning in 195 elementary schools over a four-year period. The results support a reciprocal-effects model which conceptualizes leadership within a changing, mutually-reinforcing system of organizational relationships.

Over the past several decades, a rich theoretical and empirical literature has sought to understand the contribution that leadership makes to organizational performance (Bass & Bass, 2008; Lord, 2001; Steers, 1975; Thomas, 1988; Yukl, 2002). Similarly, scholars focusing on educational settings have developed a theoretically-informed, empirical knowledge base on leadership effects (Bossert et al., 1982; Firestone & Wilson, 1985; Hallinger & Heck, 1996a, 1996b; Leithwood, Louis, Anderson, & Wahlsttom, 2004; Pitner, 1988). This research generally supports the conclusion that school leadership exerts a measurable, albeit indirect, effect on student learning (Hallinger & Heck, 1996b; Leithwood et al., 2004; Robinson, Lloyd, & Rowe, 2008). School leadership appears to achieve these effects through strategic actions and interactions that increase the school's capacity for sustained improvement through the focus on *changing* a constellation of socio-cultural, structural, and academic processes (Leithwood, Anderson, Mascall, & Strauss, in press). These processes can include the appropriateness and quality of the school's vision and goals, academic norms and expectations, allocation of resources, curricular alignment, instructional strategies, student support, as well as its previous efforts to implement targeted changes in educational practices (Firestone & Corbett, 1988; Hallinger, Bickman, & Davis, 1996; Heck, Larson & Marcoulides, 1990; Heck & Hallinger, 2009; 2010; Leithwood et al., 2004, in press; Robinson et al., 2008; Smither, London, & Richmond, 2005).

Two problems, however, complicate the interpretation of these findings for the purpose of leading and sustaining school improvement. First, most quantitative studies of school leadership effects have relied on cross-sectional data, which are ill-suited to inform on issues of change and school improvement (Bossert et al., 1982; Bridges, 1982; Hallinger & Heck, 1996a; Heck & Hallinger, 2005; Krüger, Witziers, & Sleegers, 2007). Second, researchers have shown a bias for framing leadership as the causal agent of change and driver of performance in organizations (Bridges, 1970, 1977; Heck & Hallinger, 2005). This assumption was, for example, evident in “two-factor” studies conducted during the 1980s and 1990s that examined the impact of leadership on student achievement (Hallinger & Heck, 1996a). More recent studies that conceptualize organizational variables such as staff motivation, organizational structure, and school culture as “mediating” the effects of leadership on school outcomes still share a similar assumption (Hallinger et al., 1996a; Heck et al., 1990; Leithwood & Jantzi, 1999; Marks & Printy, 2003; Wiley, 2001). Although mediated-effects studies represent a conceptual advance over earlier two-factor research, both approaches ignore the possibility that leadership *both impacts* and *is influenced by* the changing state of organizational processes and performance (Bridges, 1970, 1977; Hallinger & Heck, 1996a; Kipnis, Schmidt, Bazerman & Lewicki, 1983).The conceptualization of leadership effects as a process of mutual influence has been termed a *reciprocal-effects* model (Pitner, 1988).

In this paper, we define, measure, and examine the relationship among three organizational constructs central to our understanding of leadership and learning: collaborative leadership, capacity for educational improvement, and growth in student academic outcomes. We conceptualized these constructs as embedded within an organizational system that is dynamic rather than static and, thus, evolving over time. Scholars have noted that the causal ordering of these processes represents an important theoretical relationship with practical implications for school improvement (Hallinger & Heck, 1996a, 1996b; Heck & Hallinger, 2005; Krüger et al., 2007; Luyten, Visscher & Witziers, 2005; Ogawa & Bossert, 1995; Witziers, Bosker & Krüger, 2003). More specifically, we ask: “How are changes in collaborative leadership, school capacity for improvement, and student learning related over time?” To investigate these relationships, we develop and test a series of multilevel longitudinal models that pose the possibility of *unidirectional* and *reciprocal* effects between leadership and related elementary school improvement processes. Through testing these alternative models, we evaluate which presents a more comprehensive picture of how leadership and school improvement capacity contribute to improvement in student learning.

Overview of the Study

In this section, we define the study’s major constructs. We then present competing models describing the role that collaborative leadership plays in school improvement.

Defining the Constructs

Over the past decade there has been increasing interest in exploring the sources, means and implications of viewing school leadership beyond that which is exercised by the principal (Leithwood, Mascall, & Strauss, 2009; Ogawa & Bossert, 1995). Terms such as distributed (Gronn, 2002; Harris, 2005; Spillane, 2006), shared (Barth, 2001; Conger & Pearce, 2003; Lambert, 2002; Marks & Printy, 2003), and collaborative (Hallinger & Heck, in press) leadership all reflect a similar concern with broadening the sources, responsibilities, cohesiveness, and influence of leadership in accomplishing various types of school activities. Harris (2005) notes that within this set of school leadership models, however, there are considerable differences in the central theoretical propositions, definitions, and operationalizations of the leadership construct. The current study employs a conceptualization that we will term “collaborative” leadership.

Collaborative school leadership. Our definition of collaborative leadership focuses on strategic school-wide actions directed toward school improvement that are shared among teachers, administrators and others in the school. In the context of this study, collaborative leadership emphasizes governance structures and processes that empower staff and students, shared commitment to achieving school improvement goals, broad participation and collaboration in decision making, and shared accountability for student learning outcomes. In this paper, we do not assess the specific leadership arrangements within schools (e.g., which specific individuals work together, the cohesiveness of the leadership team). Instead, we focus on measuring teacher perceptions about the extent to which collaborative, improvement-oriented leadership exists in the school and then representing how such perceptions may change over a four-year period.

Four assumptions frame our approach to investigating how collaborative leadership may contribute to school improvement. First, the practice of leadership involves developing a shared vision for change and then enabling people to work collaboratively to achieve the vision (Bass, 1990; Hallinger et al., 1996; Leithwood et al., 2004; Yukl, 2002). Second, leadership in schools tends to be distributed among a variety of people in different roles and, therefore, its measurement should not be limited to the actions of the principal alone (Day et al., 2006; Gronn, 2002; Leithwood et al., 2009, 2009; Marks & Printy, 2003). Third, leadership should facilitate conditions that support effective teaching and learning and build capacity for professional learning and change (Fullan, 2006; Hallinger et al., 1996; Heck et al., 1990; Leithwood et al., 2004, 2009; Marks & Printy, 2003; Robinson et al., 2008; Wiley, 2001). Fourth, leadership that increases the school's capacity for improving teachers' instructional expertise will impact student outcomes positively (Fullan, 2006; Lee & Bryk, 1989; Leithwood et al., 2004; Mulford & Silins, 2003; Robinson et al., 2008; Stoll & Fink, 1996).

School improvement capacity. Increasing the school's capacity for educational improvement represents a key target of strategic leadership efforts designed to impact teacher practice and student learning (Heck & Hallinger, 2009; Leithwood et al., 2004, in press; Robinson et al., 2008). We refer to school improvement capacity as conditions that support teaching and learning, enable the professional learning of the staff, and provide a means for implementing strategic action aimed at continuous improvement (e.g., Darling Hammond, 2006; Fullan, 2001; Heck & Hallinger, 2009; Hill & Rowe, 1996; Leithwood et al., 2004, in press; Mulford & Silins, 2009; Stoll & Fink, 1996). Tapping into this underlying process with multiple measures at several points in time provides a means of monitoring the evolving organizational processes proximal to student learning across a large number of cases. We seek to develop a dynamic view of improvement in the school's educational capacity by measuring teachers' perceptions of their school's sustained efforts to improve student learning and their own classroom instruction and related instructional improvement processes. This information is then used to define a growth trajectory that portrays change in these processes for each school over the four-year period of the study.

Growth in student learning. Proponents argue that the longitudinal assessment of individual students' progress is superior to comparing successive student cohorts (e.g., percentages of students who attain proficiency each year) for the purpose of monitoring school improvement. Monitoring the progress of individual students over time captures their actual academic growth as they move through their educational careers. This focuses attention more squarely on students' experiences in attending a particular school over several years and provides a way of recognizing that schools serve students who start at different places and progress at different rates (Seltzer, Choi & Thum, 2003). We define growth in student learning as the change in math scores of a longitudinal student cohort during the study.

Defining the Models

It is only since the early 1980s that scholars began to conceptualize and study school leadership as directed toward improvement in student learning outcomes (Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger & Heck, 1996b; Leithwood et al., in press). A assumption underlying most of this research is that leadership causes change in the school and its performance (e.g., Hallinger & Heck, 1996a; Leithwood et al., 2004, 2009; Witziers et al., 2003). Theorists refer to conceptual models that reflect this assumption as *unidirectional* or *recursive models*. In a typical diagram detailing proposed relationships in unidirectional models, paths with single arrows (A → B) imply a directional effect from the first to the second variable (e.g., leadership impacts teacher commitment to change). As commonly employed in non-experimental research, we use the term *effect* to indicate a statistically significant directional relationship between the two variables (Cook, 2002).

Scholars have further noted that more recent empirical studies generally conceptualize the relationship between school leadership and school effectiveness as more indirect—that is, as mediated by school-level organizational structures and processes (Bossert et al., 1982; Hallinger & Heck, 1996b; Heck & Hallinger, 2005; Leithwood et al., 2004, in press; Pitner, 1988). Nonetheless, even these "mediated-effects" studies continue to emphasize unidirectional relationships as a means of investigating the contributions of leadership and related organizational processes. For example, in a mediated-effects model, we might conceptualize leadership as influencing teachers' commitment to change which, in turn, influences their implementation of targeted changes in instruction and student learning outcomes (A → B→ C→ D).

Overall, findings from a substantial body of mediated studies indicate that successful school leadership links the internal and external environments of the school and develops the school's capacity for change (Firestone & Corbett, 1988; Hallinger & Heck, 1996b; Heck & Hallinger, 2009; 2010; Leithwood et al., 2004, in press; Marks & Printy, 2003; Mulford & Silins, 2003, 2009; Robinson et al., 2008; Wiley 2001). Foci for school leadership in these models include developing a shared vision, increasing the academic press for learning, emphasizing teacher professional development, facilitating a collaborative working culture, and involving stakeholders in decision making (Hallinger & Heck, 1996b; Leithwood et al., 2004, in press; Robinson et al., 2008). These changes at the school level are assumed to produce "trickle down" effects on teacher classroom behavior (Creemers, 1994; Heck & Hallinger, 2009; 2010; Leithwood et al., 2004, Mulford & Silins, 2009; Robinson et al., 2008).

Defining Reciprocal-Effects Relationships

The term *leadership for learning* implies that causal linkages exist between the actions of leaders and learning outcomes. We note further that *school improvement* is, by definition, a process that involves change in the state of the organization over time. These assumptions suggest that the empirical study of school improvement leadership requires dynamic models that take into account changing relationships among relevant variables over time. As we have previously noted: “To the extent that leadership is viewed as an adaptive process rather than as a unitary independent force, the reciprocal-effects perspective takes on increased salience” (Author, 2009, p. 19).

Reciprocal-effects (or nonrecursive) conceptualizations imply two specific types of relationships beyond those encompassed in the recursive models described earlier. First, reciprocal-effects models imply the presence of *feedback loops*, or processes of mutual causation, between two variables measured at the same time. Reciprocal-effects models, therefore, suggest that two (or more) variables may be both a cause and an effect of each other (Marsh & Craven, 2006). As Jöreskog and Sörbom (1989) note, constructs such as leadership cannot directly affect themselves (A→A). Nevertheless, there may be a *total* effect exerted on each construct in a reciprocal-effects model by defining a feedback loop. As they suggest, a feedback loop is a causal chain (or cycle) from one construct passing through one or more other constructs and feeding back to the original construct (A B). One cycle consists of a path from A to B followed by a return back to A. After two cycles, the total effect will increase as the sum of a linear term (cycle one) and a quadratic term (cycle 2), and so on, provided the infinite series converges (Jöreskog & Sörbom, 1989). It should also be noted that systems should be in relative equilibrium when estimating feedback loops in order to obtain optimal estimates of effects (Kline, 2004). *Direct feedback* involves two variables in a reciprocal relationship, while *indirect feedback* involves three or more variables (Kline, 2004).

Second, a reciprocal-effects perspective further implies that earlier temporal states of component variables in a proposed model *mutually influence each other over time*. School leaders may, for example, initiate changes in work structures, management processes, curriculum content, community relations, and instructional practices. Changes in these conditions may subsequently produce effects on leadership behavior, as well as changes in distal outcomes such as student learning. Reciprocal-effects models, therefore, propose that the relationship among constructs at different points in time may be mutually reinforcing (Marsh & Craven, 2006); that is, $A_1 \rightarrow B_2$ and $B_1 \rightarrow A_2$. These temporal relationships are also referred to as “cross-lagged” relationships.

The possibility of reciprocal influence between leaders and followers has been acknowledged in the leadership literature for more than two decades (Pitner, 1988; Podsakoff, 1994). The related concepts of reciprocity, responsive adaptation, mutual influence, and leader-follower interaction are implied in theories underlying contingency leadership (Fiedler, 1967), servant leadership (Patterson, 2003; Winston, 2003), upward managerial influence (Kipnis, Schmidt, Bazerman, & Lewicki, 1983; Rao & Mawhinney, 1991; Schriesheim & Hinkin, 1990), and distributed leadership (Gronn, 2002; Spillane, 2006). Only in recent years, however, has recognition of the interactive relationship between leaders and followers led to initial empirical tests of reciprocal effects models (e.g., Griffin, 1997; Keller, 2006; Tate, 2008; Vogelaar & Kuipers, 1997).

We note that progress in testing conceptual models with reciprocal causation has been hindered by several methodological challenges. In models of reciprocal interaction there is an explicit assumption that behavioral adaptation unfolds over time (Griffin, 1997; Hallinger & Heck, 1996a; Marsh & Craven, 2006; Tate, 2008). Ogawa and Bossert (1995) succinctly summarize the case for using longitudinal data in studies that seek to examine the effects of leadership on the organization:

> [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. (239-240)

Longitudinal data are, however, difficult to obtain, especially on a scale sufficient to address questions concerning the effects of leadership across comparable organizational units. Moreover, until recently, the field lacked analytical tools capable of modeling reciprocal effects over time (Griffin, 1997; Heck & Hallinger, 2005, 2009; Marsh & Craven, 2006; Podsakoff, 1994; Tate, 2008). This problem is particularly relevant in educational organizations, where studying leadership effects on school improvement requires the use of multilevel modeling (Heck & Hallinger, 2005). Despite these challenges, however, we cannot overstate the importance of using longitudinal designs in leadership research where progress has both theoretical implications and practical utility (Heck & Hallinger, 2005; Ogawa & Bossert, 1995; Tate, 2008).

Research Focus

The primary purpose of this study is to explore the viability of a reciprocal-effects model of collaborative leadership and school improvement. As Hoyle and Panter (1995) noted, in practice researchers may not have one conceptual model in mind but, rather, a series of competing models. Testing the adequacy of several proposed models in sequence is known as an "alternative models" approach. Testing alternative models is often useful in circumstances where the research seeks to compare the adequacy of established models against new conceptual formulations.

In this study we investigate three models that propose contrasting theoretical relationships among our major constructs (see Figure 1). Because we focus on comparing alternative sets of relationships, for simplification of the presentation, we only show the contrasting *between-schools* portion of the models. Each proposed model also incorporates a *within-schools* model; this specifies a set of student background variables proposed to explain variation in individual growth in mathematics over time.

We used latent change analysis (LCA), a type of structural equation modeling (SEM) used for investigating longitudinal data, to test our proposed models. In LCA, changes in individual and organizational behavior are conceptualized as continuous latent *level* (intercept) and *change* (or growth) factors (Raykov & Marcoulides, 2006). In this type of model, the factor scores describing the constructs are measured on several occasions. One advantage of the SEM approach is that several developmental processes (with direct, indirect, or reciprocal effects) can be represented simultaneously within a single model. This enables optimal estimation and direct incorporation of measurement error in the analysis. This makes the method appropriate for modeling complex and changing organizational processes such as leadership and school improvement.

The first two models summarized in Figure 1 provide a reduced set of unidirectional effects between the level factors (i.e., unshaded ovals that represent the initial status of each construct) and the change factors (i.e., shaded ovals that represent the change occurring between successive measurements). We note the models also specify a correlation between each initial status and its corresponding change factor (A←→ B). Correlations imply no causal relationship. It is common in growth models for the initial status and growth factors to be negatively correlated. For example, due to any number of possible reasons (e.g., less perceived need to change, ceiling effects in measuring achievement), we generally expect that schools with high initial achievement levels will grow less in achievement over time than schools that start with low initial levels of achievement. In contrast, if we examined *ending* achievement, the correlation between the change and ending achievement factors would be positive.

The third model represents a complete reciprocal-effects system. The three *initial status* factors are proposed to be mutually reinforcing over time (i.e., specified through a set of cross-lagged path relationships from each initial status factor to a different change factor), and also the model is proposed to incorporate an indirect feedback loop between the *change* factors. The indirect feedback loop represents the extent to which initial growth (or change) rates may also explain subsequent growth rates over time.

Insert Figure 1 about here

Model 1: Does Change in Leadership Affect School Growth Indirectly?

The first model proposes that change in collaborative leadership is a driver for change in school improvement and subsequently in school learning outcomes. Its most important feature is an implied *indirect* positive relationship between leadership and growth in student learning through building school improvement capacity. This model includes only *unidirectional* relationships. Support for this model would be evident if we found significant *direct* effects between the change constructs, as well as a significant *indirect* effect between leadership and growth in student achievement.

Model 2: Does Change in School Outcomes Affect Change In Leadership Indirectly?

The second model is also unidirectional, but the direction of the recursive effects operates in the opposite direction from Model 1. This model investigates the possibility that school growth in math may influence school improvement capacity and, indirectly, collaborative leadership. Support for this model would be provided if we found significant *direct* effects of school growth in math on school improvement capacity and if improvement capacity positively influenced changes in collaborative leadership. Support would also be strengthened if school growth in learning *indirectly* affected change in school leadership through change in school instructional capacity.

Model 3: Do the Processes Represent a Mutually-Reinforcing System?

Model 3 represents an attempt to conceptualize the relevant variables as a set of reciprocal, or mutually reinforcing, processes. In a full nonrecursive path model, each endogenous (i.e., mediating and dependent) variable is directly affected by all other endogenous and exogenous (independent) variables in the model (Berry, 1984). As a practical matter, however, this type of fully-specified nonrecursive model is difficult (and sometimes impossible) to fit to the observed covariance matrix. Analysts usually eliminate some possible paths in order for the proposed relationships to converge on a solution. Model 3 concerns whether the initial state of each construct and subsequent changes in the other two constructs are mutually reinforcing over time. Moreover, it proposes that change rates in these constructs are related to each other through an indirect feedback loop. These features of the model therefore suggest that changes in the three organizational constructs are both cause and effect of each other (Kline, 2004).

This model would be supported if three conditions were met. First, unidirectional effects from the initial status factors to the change factors should be statistically significant. Second, the indirect feedback loop between change in leadership, change in school improvement capacity, and school growth in learning should all be significant and substantial in size. Third, it should provide the best fit to the data among the models.

Method

Sample and Instrumentation

A sample of 195 elementary schools was randomly selected from the population of elementary schools in a western state in the United States. Within those schools, a longitudinal cohort consisting of all third-grade students within the schools (N = 13,391) participated in the study (*M* = 92.18, *SD* = 43.57). Student demographics are shown in Table 1. Student socioeconomic status (SES) was estimated by determining participation in the state's federally-funded lunch program (45% of the students in the study participated in this program). Fourteen percent of the students entered the school system after the first year of the study, and 16 percent changed schools. One advantage of the LCA approach to modeling change 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).

The study utilized longitudinal state-wide survey data on leadership and school improvement capacity from teachers on three occasions over a four-year period to compare the fit of three multilevel models. These surveys are regularly administered to all certificated staff, all grade five students, and a random sample of parents (i.e., approximately 20% across grade levels) with the goal of monitoring the quality of key educational processes (e.g., collaborative leadership, implemented of standards-based instruction, professional capacity, focus on academic improvement) over time. 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 [i.e., often referred to as a longitudinal panel study (Davies, 1994)]. 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 capture potential changes in leadership and academic processes using the surveys given to each school's teachers on three occasions (i.e., years 1, 3, and 4).[i] Total teacher return rates for the three periods of data collection in this study were 73.4%, 76.4%, and 75.6%, respectively. School return rates ranged from 47% to 100% (Mean = 74.4%, SD = 6.2%) across the period of data collection.

Data on individual student math outcomes were collected in years two, three, and four. At each measurement occasion, the data from school surveys precede the student achievement data. Unequal spacing of observations and nonlinearity can be incorporated into an LCA model without compromising the quality of data analysis (Raykov & Marcoulides, 2006).

The surveys have undergone pilot testing and subtle changes in the items comprising each dimension of schools' educational processes over the 12 years that they have been used to monitor opportunity-to-learn conditions in schools. Both face and predictive validity of the individual items comprising the scales used to define the leadership and school improvement capacity constructs has been assessed previously. For example, prior studies found a significant relationship between the quality of schools' academic capacity (e.g., educational expectations, academic press for achievement, curricular and instructional processes, communication, student well being) measured by the school surveys and elementary school students' achievement in math (**add author citations*). 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.

Operational Definition and Measurement of Variables

Student background and school context controls. Student background controls included gender, ethnicity, socioeconomic status, special education status, language background, and student mobility. School context indicators describe initial school contexts during the first year of the study, unless otherwise noted. In preliminary analyses we investigated a number of school input and process variables (e.g., enrollment size, staff characteristics including average teacher experience, staff and principal stability, teacher professional preparation, student composition, school academic and social organization). Although some of the input variables were found to influence either initial achievement levels or growth independently, as a set, school inputs did not appreciably alter the relationships among our three primary constructs.

We did, however, retain two variables in our final analyses—student composition (which influenced initial leadership, initial capacity, and school growth consistently and negatively) and principal stability (which positively influenced perceptions of change in school improvement capacity over time). *Student composition* was a composite variable developed by combining several student demographics to create a weighted indicator using principal components analysis ($M = 0$, $SD = 1$). The variables included the percentage of children receiving free or reduced lunch, percentage of students receiving English language services, and the percentage of minority students. Larger positive values represent school settings where these percentages of students were higher. *Principal stability* assessed if the same principal (coded 1, else = 0) was at the school during the four years of the study.

Collaborative leadership. Items used to collaborative leadership were measured on five-point, Likert-type scales and expressed as the percentage of positive agreement with each statement. The items used to describe collaborative leadership were preliminarily developed using multilevel (teachers nested within schools) confirmatory factor analysis (CFA) and a sample of over 4,056 teachers in a subset of 51 schools. At the teacher level, item loadings were all statistically significant and ranged from 0.35 to 0.99. At the school level, item loadings were all above 0.93. Cronbach's alpha (α), a measure of internal subject consistency, provides another indicator of reliability.

Collaborative leadership ($\alpha = 0.85$) was measured by eight items which are consist with our four preliminary assumptions about collaborative leadership (i.e., emphasizes shared vision and collaborative work; distributed among role groups in the school; focused on building conditions for professional learning and change; directed toward improvement of conditions supporting learning). The items measure three specific aspects of collaborative leadership (with items paraphrased in parentheses):

- *Shared school governance* which encourages shared vision, broad participation, and shared accountability for student learning (e.g., Teachers, administrators, and staff work together effectively to achieve school goals; Teachers can freely provide input and express concerns to administrators; School provides opportunities for parents to participate in important decisions about their children's education (e.g., scheduling, homework, discipline);
- *Collaborative decisions* focusing on academic improvement (e.g., School ensures teachers have a major role in decisions about standards-based curriculum development; School provides opportunities for teachers to plan and make school decisions about professional development and curriculum; Teachers have needed instructional resources to teach effectively;
- *Broad participation* in efforts to evaluate the school's academic development (e.g., School provides regular opportunities for all stakeholders to review the school's vision and purpose; Assesses progress in making school changes).

School instructional capacity. We operationalized school improvement capacity from four subscales describing the extent to which the school has educational programs that are aligned to state curriculum standards, seeks ways to implement programs that promote student achievement over time, has a well-developed range of academic and social support services for students, and has a professional teaching staff well qualified for assignments and responsibilities and committed to the school's purpose. Each subscale was preliminarily defined from multiple survey items collected from the subset of teachers within 51 schools as using CFA as described previously (factor loadings on each subscale were all statistically significant and substantial in size). The four subscale alphas and items of the subscales are next described.

The first subscale was *standards emphasis and implementation* (α = 0.91). Items included the following: School's educational programs are aligned to state content and performance standards; Teaching and learning activities are focused on helping students meet 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.

The second subscale was *focused and sustained action on improvement* (α = 0.82). Items included the following: 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.

The third subscale was *quality of student support* (α = 0.85). Items included the following: 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.

The fourth subscale was *professional capacity of the school* (α = 0.80). Items included the following: 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.

After developing our leadership and school improvement capacity subscales, we used the multiple-group capacity of the Mplus software program and CFA to test the fit of the set of subscales to their respective leadership and school improvement capacity factors across the three measurement occasions. This is known as testing the measurement invariance of the observed indicators defining the factors and is a necessary step in the investigation of underlying changes in organizational processes over time (Raykov & Marcoulides, 2006). We then saved the three sets of collaborative leadership and school improvement capacity factor scores in the data base for the subsequent analyses of change (see Table 1). Positive scores suggest greater relative change over time compared with the initial status means (i.e., which are fixed to 0.00).

Math outcomes. Math scores for the cohort of students in the study were collected over three successive years (third-fifth grades). The math tests were constructed in relation to state-developed curricular goals. The tests consisted of constructed-response items and standardized test items from the Stanford Achievement Test (Edition 9). For math, there were five curricular strands corresponding to state learning standards consisting of 52 items (i.e., number and operation; measurement; geometry and spatial sense; patterns, functions and algebra; data analysis, statistics, and probability). Resulting math scaled scores (which were equated to range from 100 to 500) considered patterns of right, wrong, and omitted responses and item difficulty over successive years.

Descriptive Statistics for Variables in the Models

Descriptive statistics for the variables are provided in Table 1. Intraclass correlations, which describe the variance in student math achievement attributable to differences between schools, ranged from 14.8 percent to 15.3 percent, consistent with previous multilevel research on school effects (Hill & Rowe, 1996). Within schools, student math scores between third and fifth grades increased by about 33.7 scaled score points in math. Student gains were slightly larger between grades 3 and 4 than between grades 4 and 5. Between schools, students' average yearly growth rate in math was 16.5 points (*SD* = 15.5). The standard deviations associated with the observed indicators and minimum and maximum percentages in Table 1 suggest considerable variability between schools in teachers' perceptions regarding the variables used to define initial levels of collaborative leadership and school improvement capacity.

Insert Table 1 About Here

Data Analysis

Specifying the models. In LCA, repeated observations over time can be expressed as factor loadings on continuous latent, or underlying, factors (Raykov & Marcoulides, 2006). The *level* factor represents the mean of the factor at a chosen point in time. In this case, we defined the level factors to indicate the initial status of each construct. The leadership and school improvement capacity initial status factors were measured in year 1 and again during year 3 and year 4. The math initial status factor was measured in years 2, 3, and 4. The *shape* factor represents the growth, or change, over a chosen interval. In the proposed models, we defined the *change* factors to incorporate possible nonlinearity in each developmental trajectory between the *second* and *third* measurement occasion. In such non-linear formulations, the path coefficient is interpreted as describing the rate change occurring between the *first* and *second* occasion. We provide additional information about the specification of the models as an endnote.[ii]

In specifying and testing more complex models with reciprocal effects, it is important to make sure we have met model identification rules (i.e., each latent variable is assigned a scale of measurement, the number of free parameters estimated must be less than or equal to the number of nonredundant elements in the observed covariance matrix, every latent variable with unrestricted variance must emit at least two directed paths to observed indicators or other latent variables when the latter variables have unrestricted error variances), as summarized by Bollen and Davis (2009). Because full nonrecursive models can result in too many paths between variables to be estimated, given the characteristics of the proposed model and sample covariance matrix summarizing the data (a problem known as *under-identification*), it is usually necessary to fix some possible relationships in the Beta matrix describing structural relationships between constructs to 0.[iii] These restrictions generally facilitate the estimation of a proposed model that converges on a unique solution.

Although the presence of feedback loops (i.e., reciprocal causation) may be a good compromise for variables measured at the same time, such models can be challenging to fit to the data and interpret.[iv] Investigating feedback loops also requires the assumption of equilibrium; that is, their estimation does not depend on the particular time in which the data were collected (Kline, 2004). We investigated this latter possibility by examining a series of preliminary cross-lagged models, each consisting of two waves of data. Such models propose that an earlier state of A affects a later state of B and vice versa (i.e., $A_1 \rightarrow B_2$ and $B_1 \rightarrow A_2$ and so on). Such models provide a alternative way of specifying mutually-reinforcing relationships (Marsh & Craven, 2006) from our LCA approach. We found the results similar over each subset of measurement occasions tested. These preliminary results support for the assumption that the structure of the proposed reciprocal relationships in Model 3 was stable over time.

Model comparisons. We used Mplus 5.21 (Muthén & Muthén, 1998-2006) to conduct the tests of our proposed models. Comparison tests are conducted by proposing a baseline model and then using several model fit indices to evaluate how well alternative models fit the data compared with the baseline model. In this way, we can test our specific hypotheses about the direction of key parameters in our proposed models.

First, we examined changes in the school improvement capacity and collaborative leadership factor scores over time. We used the multiple-group capacity of SEM and confirmatory factor analysis to test the fit of the subscales to the factors across the three measurement occasions (Raykov & Marcoulides, 2006). We conducted this analysis to establish the reliability and validity of our conceptualization of collaborative leadership and school improvement capacity over several measurement occasions. We wanted to establish whether the constructs were measured consistently over time and the extent to which schools improved in their capacity to provide collaborative leadership and quality educational practices over the period under study. To examine whether mean teacher perceptions did in fact change, successive factor means in Table 1 can be tested with *t*-tests against the initial factor means (0.00). These tests suggested collaborative leadership and improvement capacity factor means were not the same over time ($p < .05$, not tabled).

Second, the analyses of the alternative models specified in Figure 1 involve imposing sets of model restrictions on the sample covariance matrix and determining which set fits best within the population under study. The adequacy of each model can be determined through the comparison of several model-fit criteria (see Marcoulides & Hershberger, 1997), as well as the size and significance of the standardized path coefficients. The chi-square coefficient provides an often-used estimate of the model's fit (i.e., with smaller values indicating better fit), but it can be inflated by sample size. The Comparative Fit Index (CFI) evaluates the fit of the proposed model against a type of baseline (non-fitting) model. Values near 0.95 provide evidence of an adequate model fit. The standardized root-mean-square residual (SRMR) coefficient describes the average magnitude of model residuals. Values of 0.05 or lower generally indicate an adequate fit of the model to the data. We report the SRMR only for the between-schools portion of the model, since the within-groups SRMR was small for all models (.003). Related to the

SRMR, the root-mean-square error of approximation (RMSEA) provides another estimate of discrepancy in the model's fit, with values near 0.05 also indicating an acceptable fit. Finally, keeping in mind that initial math achievement scores (if included in each model) would account for most of the between-school variance in math growth rates, we can estimate the variance accounted for (R^2) in school growth rates from the set of relationships in each model.

Results

Our first concern is the fit of each of the models proposed in Figure 1 to the data. If the models do not fit the data adequately, we would have to reconceptualize them. It is important to keep in mind, however, that *failure to reject* a particular model does not mean that it is the only correct model. Other models also might fail to be rejected on statistical grounds. This is why we examined the robustness of each proposed model using multiple criteria. Table 2 provides a summary of several model-fit indices.

Overall, each of the three models fit the data adequately when judged against the specified criteria (e.g., CFI values above 0.95, SRMR values equal to or below 0.05). When compared against each other, however, Table 2 suggests that the reciprocal-effects model (Model 3) provided the strongest fit to the data. More specifically, the chi-square coefficient was the lowest among the three models, and the CFI, SRMR, and RMSEA fit indices all supported this conclusion. According to the chi-square coefficient, Model 1 fit the data better than Model 2, for the same degrees of freedom. We can also test the difference in chi-square coefficients between Model 1 versus Model 3, for 6 degrees of freedom and a scaling correction applied for unequal sample sizes in multilevel models (Muthén & Muthén, 1998-2006). Table 3 indicates $\Delta\chi^2$ is larger (14. 48) than the required coefficient of 12.51 ($p < .05$). This suggests Model 3 provided an improved fit to the data over Model 1 (or Model 2). Relationships implied in Model 3 also accounted for more variance in school growth rates ($R^2 = 0.18$) than the other two models.

Insert Table 2 about Here

Examining the Parameter Estimates for Each Model

Although Model 3 provided the strongest statistical fit to the data, all three proposed models fit the data adequately. Therefore, it is incumbent upon us to compare the viability of each model according to its parameter estimates further. Figure 2 presents a comparison of the between-school standardized estimates for each model. These can be interpreted as effect sizes since they are all in the same metric. The within-school standardized estimates for student background (e.g., gender, SES) and the between-school context controls (e.g., student composition, principal stability) were the same in each model.[v] The figure also indicates that initial perceptions about collaborative leadership in each model were unrelated to perceptions about initial school improvement capacity, implying at the beginning of the study these constructs were independent.

Turning to our proposed change relationships, Model 1 suggested that change in collaborative leadership was positively related to change in school improvement capacity (0.37, $p < .05$); in turn, change in school improvement capacity positively affected school growth in math (0.24, $p < .05$). Moreover, there was a significant indirect effect of change in leadership on math growth rates (0.09, $p < .05$; not tabled).

Insert Figure 2 about Here

Model 2 provided little empirical support for the premise that school growth in math learning influenced perceptions of changes occurring in school improvement capacity. Change in school improvement capacity, however, was predictive of perceptions of change in collaborative leadership. Because only one proposed change relationship was statistically significant, there was also no indirect relationship between school growth rates and perceptions of changes in leadership, as implied in this model.

Model 3 provided considerable evidence in support of our reciprocal-effects perspective on leadership and school improvement. First, we found that initial achievement was positively related to subsequent changes in both collaborative leadership (0.12, $p < .05$) and change in school improvement capacity (0.33, $p < .05$). Importantly, however, the converse was not true—that is, neither initial levels of leadership nor initial levels of school improvement capacity were directly predictive of subsequent growth rates in math. This provides empirical support for the premise that schools can improve outcomes *regardless* of their initial achievement levels by changing key organizational processes (i.e., leadership and school improvement capacity).

Second, initial school improvement capacity was related to subsequent change in leadership (0.39, $p < .05$), and initial collaborative leadership was related to subsequent changes in school improvement capacity (0.18, $p < .05$). This result implies that leadership and school improvement capacity represent mutually-reinforcing, or parallel, change processes--each *initial status* factor explaining positive growth in the other *change* factor. The initial effect of improvement capacity on subsequent leadership change was about *twice* as strong as the effect of initial collaborative leadership on subsequent change in school improvement capacity.

Third, the presence of the hypothesized indirect feedback loop was confirmed. More specifically, change in collaborative leadership was related positively to change in school improvement capacity (0.17, $p < .05$), and change in improvement capacity was related positively to student growth rate (0.19, $p < .05$). Conversely, school growth rate was also predictive of change in school improvement capacity (0.24, $p < .05$), and change in school improvement capacity was predictive of change in leadership (0.37, $p < .05$). This finding is also consistent with our premise that change in schools' outcomes can be the impetus for further changes in leadership and other organizational processes.

Model 3 therefore supported the proposition that changes in collaborative leadership and school improvement capacity are mutually-reinforcing processes—that is, changes in the organization "gain momentum" over time through changes in leadership and school improvement capacity that are organic and mutually responsive. In addition, the indirect feedback loop suggests that over time the total effects of organizational leadership can increase as a function of changes in both improvement capacity and school growth. Importantly, for interpreting the meaning of the indirect feedback loop, we determined through our preliminary cross-lagged analyses of leadership and improvement capacity that the difference in standardized effects between the two change factors (0.17 and 0.37) in Figure 2 implies that the effect of school improvement capacity on collaborative leadership was *stronger* at each measurement occasion than the corresponding effect of leadership on improvement capacity. In contrast, reciprocal coefficients for improvement capacity and achievement (0.19, 0.24) imply the effects on each other were similar in size over time. This additional empirical evidence would remain hidden in models containing only unidirectional-effects.

Discussion and Implications

The purpose of the study was to examine the changing relationships among school leadership, school improvement capacity, and student math achievement over time. Three features distinguish this study from prior research on school leadership effects. First, the study conceptualized leadership as a collaborative process aimed at building the school's academic improvement (Barth, 2001; Conger & Pierce, 2003; Fullan, 2001; Gronn, 2002). Few studies conducted to date have examined the impact of collaborative leadership in educational organizations (Leithwood et al., 2009; Marks & Printy, 2003).

Second, this research responded to calls for studies that utilize longitudinal data to explore leadership effects on organizational processes and performance (Bridges, 1982; Griffin, 1997; Hallinger & Heck, 1996a; Heck & Hallinger, 2005; Ogawa & Bossert, 1995; Pitner, 1988; Tate, 2008). This was one of the first large-scale studies examining relationships among school leadership, capacity-building processes, and outcomes using longitudinal data (see also Mulford & Silins, 2009). This enabled us to offer some preliminary insights into how the relationship between leadership and organizational improvement evolves over time (Ogawa & Bossert, 1995; Tate, 2008).

Third, the study represented an initial empirical effort to evaluate the usefulness of reciprocal-effects models in studies of leadership and school improvement. Indeed, our search of the broader literature identified only two other reciprocal-effects studies of leadership (Griffin, 1997; Tate, 2008). Reciprocal-effects formulations offer new concepts for describing relationships in leadership research—in particular, mutually reinforcing, or parallel change, processes and feedback loops. In this section we draw several conclusions, highlight limitations, and provide some thoughts regarding how this empirical effort furthers our understanding of leadership and school improvement.

Conclusions

This research explored alternative ways of specifying the relationship between collaborative leadership and school improvement. Although studies of school leadership effects conducted over the past 20 years have tended to employ mediated-effects models (Heck & Hallinger, 2005; Leithwood et al., 2004, in press), scholars have argued for the need to explore the possibility that leadership effects on relevant school improvement processes are reciprocal or mutually reinforcing (Edvantia, 2005; Hallinger & Heck, 1996b; Heck & Hallinger, 2009; Krüger et al., 2007; Pitner, 1988; Southworth, 2002). In this paper, we formulated several alternative conceptual models designed to explore our central research question concerning how changes in collaborative leadership, school capacity for improvement, and student learning are related over time.

Path relationships in Model 1 proposed that collaborative leadership impacts school instructional capacity, which in turn leads to direct (and indirect) changes in student learning outcomes. This mediated-effects model was generally supported by the data. In contrast, Model 2 proposed that when a school's achievement improves or declines over time, some predictable patterns of change in perceptions about improvement capacity (e.g., curriculum alignment, instructional strategies) and leadership will occur in response. This model posited *results* as the driver of changes in the school's capacity for improvement, and collaborative leadership as responsive to these changes. Although Model 2 fit the data adequately, it was only partially supported from a substantive standpoint. We therefore concluded that these two unidirectional models offered only partial understandings of how leadership effects unfolded over time, even though the hypothesized *direct effects* of leadership on school improvement capacity and *indirect effects* on growth in student achievement were clearly evident in Model 1.

Model 3 employed a reciprocal-effects perspective. In testing this model, we posited the more specific question: “Do initial organizational states of leadership, school instructional capacity, and student achievement affect subsequent changes in these variables?” The results yielded the following conclusions:

- The reciprocal-effects model provided the strongest fit to the data, based on model-fit criteria, and the majority of proposed substantive relationships were statistically significant.
- Initial achievement was positively related to subsequent changes in school improvement capacity and collaborative leadership.
- Collaborative leadership and school improvement capacity evidenced reciprocity, or mutual reinforcement; that is, each initial-status construct explained subsequent changes in the other growth construct. This relationship positively influenced school growth in math achievement over time.
- Changes in collaborative leadership compounded over time through an indirect feedback loop consisting of changes in school improvement capacity and growth in student math achievement. Moreover, leadership effects on subsequent school improvement capacity were *smaller* than corresponding capacity-building effects on subsequent collaborative leadership.

In concert with these findings, we suggest that Model 3 is the most theoretically compelling of the three models. First, it does not make untenable assumptions about the heroic role of leadership and presents leadership for learning in dynamic relationship with other organizational processes. Second, our results offer a new empirical description of collaborative leadership and school educational capacity building as mutually reinforcing, parallel change processes. More specifically, our results imply that collaborative leadership was not only an initial driver of change in school improvement capacity, but it also subsequently mediated the initial effects of school improvement capacity and initial school achievement on subsequent changes in improvement capacity and school growth in math. Third, the presence of an indirect feedback loop supports the premise that the total effects collaborative leadership, as an organizational property (Ogawa & Bossert, 1995) can increase (or decrease) over time as a function of changes in schools' evolving educational capacity building efforts and achievement trajectories. Thus, this model captures the dynamic and responsive nature of leadership for learning.

More complex theoretical conceptualizations require the use of causal inference techniques that can handle a wider variety of theoretical relationships in multilevel organizational settings (Griffin, 1997; Kline, 2004; Marsh & Craven, 2006). We were encouraged by the results of our various model tests. Multilevel SEM can be a useful tool in defining and testing longitudinal models with reciprocal effects; however, we caution that the results should be considered carefully within the context of several guidelines. These include the theoretical foundation underlying the model tested, the reliability and validity of the measured variables used to define the constructs, the nature of the relationships between the latent variables (e.g., direction of causality), and the plausibility of the results (e.g., the fit of the model to the data). Our initial examination of the measurement model used to define the latent collaborative leadership and improvement capacity factors within and between schools and at three points in time suggested that they can be reliably measured. We next provided evidence that teachers' perceptions about schools' instructional capacity and collaborative leadership changed over time. Only then did we test the suitability of our proposed conceptual models to the data.

The reciprocal-effects model supported in this paper provides a basis for exploring the inclusion of multiple growth trajectories, mutual influence, and compounding effects when examining leadership and school improvement. In sum, the results suggest that the reciprocal-effects model offers a more comprehensive view of how leadership interacts with other processes in dynamic organizational systems.

Limitations

We also note a number of limitations of this research. First, questions remain about the day-to-day development of leadership that is exercised collaboratively and is aimed at improving school improvement capacity. School-level questionnaires are an imperfect means of extracting information about organizational processes. For example, although the existing items measured leadership as a collaborative process reliably, an individual's reported involvement in school decision making may, or may not, adequately capture a key aspect of collaborative leadership. Moreover, the available data did not support the up-close description of ways in which collaborative leadership (and various leadership arrangements to exercise it) may have differed across various school contexts.

Second, this study focused on three primary constructs (i.e., collaborative leadership, school improvement capacity, and math achievement) consisting of multiple items measured on multiple occasions. There are likely other important constructs that may also operate within the causal system, thereby leaving our model open to potential misspecification (Cook, 2002; Griffin, 1997). For example, although our research adopted a multilevel model, the study did not use observed measures of teachers' classroom practice. Thus, the causal link between school improvement capacity, changes in classroom practices, and student achievement remains a "black box." We acknowledge that the type of school-level aggregates employed in this study can ignore wide variations in teaching and learning conditions at the classroom level which are known to affect student outcomes (McCaffrey, Lockwood, Louis, & Hamilton, 2004).

Third, it is also important to acknowledge that such models may still not resolve issues of whether A causes B or B causes A—unless the assumptions previously outlined can be met. As we noted, one assumption is the stability of the model's causal structure, that is, that the structure of proposed relationships should not substantially change over time. In our preliminary analyses, we fit models with various combinations of two waves of data and found the results to be similar to our final models. These tests using a slightly different modeling approach provided support for the assumption that the structure of the proposed relationships was the same over the three waves of data collected. It may be the case, however, that longer periods of time are needed for effects to be fully observed.

Implications

Despite these limitations, we hope that the initial support offered by this empirical study for viewing leadership effects as a reciprocal process will stimulate scholars to engage in new theory building about leadership and school improvement. Researchers have become increasingly comfortable with the view that leadership effects on student learning are indirect in nature, yet this perspective still requires an uncomfortable assumption about the role of leadership in complex organizations (Bossert et al., 1982; Bridges, 1970, 1977, 1982). Although leadership may well function as a catalyst for change at certain points in time, this perspective fails to account fully for its interactions with other supporting and constraining forces in the organization and its environment.

According to a reciprocal-effects conceptualization of school improvement leadership, analysts would carefully monitor data feedback on changes in several parts of the system simultaneously. The results thus portray the school improvement process as one in which the organization "gains momentum" over time through internal changes that are organic and mutually responsive. In turn, improvements in learning also loop back to stimulate the development of more refined instructional capacity-building activities and leadership strategies.

With this latter point in mind, we assert the need to reconsider the usefulness of contingency theories of leadership. Research needs to examine how leadership responds to changing environmental forces as well as to changes in organizational processes and performance outcomes over time. Given current theoretical constructs and statistical methods, conducting further reciprocal-effects studies with longitudinal data is eminently possible and highly recommended.

References

Barth, R. (2001). Teacher Leader. *Phi Delta Kappan*, 443-449.

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

Bass, B., & Bass, R. (2008). *The Bass handbook of leadership*. New York: Free Press.

Berry, W. D. (1984). *Nonrecursive causal models*. Thousand Oaks, CA: Sage.

Bollen, K. A. & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling, 16*, 498-522.

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

Bridges, E. (1970). Administrative man: Origin or pawn in decision making? *Educational Administration Quarterly, 6* (1), 7-25.

Bridges, E. (1977). The nature of leadership. In L. Cunningham, W. Hack, & R. Nystrand (Eds.), *Educational administration: The developing decades.* Berkeley: McCutchan.

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

Conger, J., & Pearce, C. (2003). A landscape of opportunities: Future research on shared leadership. In C. L. Pearce and 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? E*ducational Evaluation and Policy Analysis, 24*(3), 175-200.

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

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

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

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

Edvantia, (2005). *Shared leadership and school improvement.* Research brief published by the Appalachia Educational Laboratory, Nashville, TN.

Fiedler, F. (1967). *A theory of leadership effectiveness.* New York: McGraw-Hill.

Firestone, W., Corbett, D. (1988). Planned organizational change. In N. J. Boyan (Ed.) *Handbook of Research on Educational Administration* (pp. 321-341). New York: Longman.

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

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

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

Griffin, M. A. (1997). Interaction between individuals and situations: Using HLM procedures to estimate reciprocal relationships. *Journal of Management*, *23*(6), 759-773.

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

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. H. (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. H. (1996b). Reassessing the principal's role in school effectiveness: A review of the empirical research, 1980-1995. *Educational Administration Quarterly, 32*(1), 5-44.

Harris, A. (2005). Leading or misleading? Distributed leadership and school improvement. *Journal of Curriculum Studies, 37*(3), 255-265.

Hayduk, L. (2009). Finite feedback cycling in structural equation models. *Structural Equation Modeling, 16*(4), 658-675.

Heck, R. H. & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement**.** *American Educational Research Journal, 46*, 626-658.

Heck, R. H. & Hallinger, P. (2010). Examining the moderating effect of instructionally-focused leadership on teacher effectiveness and student learning. Paper presented at the annual meeting of the American Education Research Association, Denver, May.

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

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

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

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL7 User's Guide*. Mooresville, IN: Scientific Software.

Keller, R. T. (2006). Transformational leadership, initiating structure, and substitutes for leadership: A longitudinal study of research project team performance. *Journal of Applied Psychology, 91*(1), 202-210.

Kipnis, D., Schmidt, D.S., Bazerman, M., & Lewicki, R. (1983). *Negotiating in organizations.* Beverly Hills, CA: Sage.

Kline, R. B. (2004). *Principles and practices of structural equation modeling* (2nd Edition). New York: Guilford Press.

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

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

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

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

Leithwood, K. & Jantzi, D. (1999). The relative effects of principal and teachers sources of leadership on student engagement with school. *Educational Administration Quarterly, 35*, 679-706.

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

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.

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

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

Marcoulides, G.A. & Hershberger, S. (1997). *Multivariate statistical methods*. Mahwah, NJ: Lawrence Erlbaum.

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

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

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

Mulford, B., & Silins, H. (2009). Revised models and conceptualization of successful school principalship in Tasmania. In B. Mulford and B. Edmunds (Eds.) *Successful school principalship in Tasmania.* Launceston, Tasmania: Faculty of Education, University of Tasmania.

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

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

Patterson, K. A. (2003). *Servant leadership: A theoretical model*. Dissertation Abstracts International, *64*(02), 570. (UMI No. 3082719)

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

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

Podsakoff, P. (1994). Quantitative methods in leadership research. *Leadership Quarterly*, *5*, 1-2.

Rao, R. K., & Mawhinney, T. C. (1991). Superior-subordinate dyads: Dependence of leader effectiveness on mutual reinforcement. *Journal of the Experimental Analysis of Behavior, 56,* 105-118.

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

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

Schriesheim, C. A., & Hinkin, T.R. (1990). Influence tactics used by subordinates: A theoretical and empirical analysis and refinement of the Kipnis, Schmidt, and Wilkinson subscales. *Journal of Applied Psychology. 75*(3), 246-257.

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

Smither, J. W., London, M., & Richmond, K. R. (2005). The relationship between leaders' personality and their reactions to and use of feedback: A longitudinal study. *Group & Organization Management, 30*(2), 181-210.

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

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

Stoll, L., & Fink, D. (1996). *Changing our schools: Linking school effectiveness and school improvement*. Buckingham: Open University Press.

Steers, R. (1975). Problems in the measurement of organizational effectiveness. *Administrative Science Quarterly*, *20*(4), 546-58.

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.

Thomas, A. (1988). Does leadership make a difference to organizational performance? *Administrative Science Quarterly*, *33*(22), 401-17.

Vogelaar, L. W., & Kuipers, H. (1997). Reciprocal longitudinal relations between leader and follower effectiveness. *Military Psychology, 9,* 199-208.

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

Winston, B. E. (2003). *Extending Patterson's servant leadership model: Explaining how leaders and followers interact in a circular model*. Paper presented at the 2003 meeting of the Servant Leadership Research Roundtable, Virginia Beach, VA.

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. (2002). *Leadership in organizations*. Upper Saddle River, NJ: Prentice Hall.

TABLE 1
Descriptive Statistics for Variables in the Models

Variable	Mean	SD	Minimum	Maximum
Within Schools (N =13,391)				
Student Outcomes				
Math 2004 (ICC[a] = 14.8%)	217.12	63.43	100.00	500.00
Math 2005 (ICC = 15.1%)	235.00	68.42	100.00	500.00
Math 2006 (ICC = 15.3%)	250.83	66.32	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
Minority	0.50	na	0.00	1.00
Changed Schools	0.16	na	0.00	1.00
Entered Year2/3	0.14	na	0.00	1.00
Between Schools (N = 195)				
School Achievement				
Initial Math Level	247.45	5.86	108.33	294.09
Math Growth Rate	16.52	5.51	-11.91	76.52
Context				
ELL (%)	8.45	9.02	0.00	61.00
Low SES (%)	50.49	22.63	0.00	97.00
Minority Mean (%)	51.16	23.97	3.00	97.00
School Composition Factor	-0.03	0.95	-1.98	2.31
Same Principal	0.31	na	0.00	1.00
Initial Collaborative 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

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

TABLE 1 (con.)

Variable	Mean	SD	Minimum	Maximum
Initial School improvement capacity				
Learning Standards (%)	87.10	6.32	69.14	98.67
Student Support (%)	78.48	10.93	37.63	98.77
Professional Capacity (%)	74.53	11.82	40.01	99.11
Focused Improvement (%)	78.41	11.26	47.22	97.35
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

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

TABLE 2
Comparisons of Model Fit Indices

	Model 1	Model 2	Model 3
Chi-Square Coefficient	161.01	198.96	145.96
Degrees of Freedom	49	49	43
Comparative Fit Index	0.99	0.99	0.99
RMSEA	0.01	0.02	0.01
Between Groups SRMR	0.05	0.08	0.03
$\Delta\chi^2$ M_1-M_3 (6 df)			14.48*
R^2 (Math Growth)	0.09	0.05	0.19

*$p < .05$

Figure 1
Proposed Alternative Models of Collaborative Leadership and Change

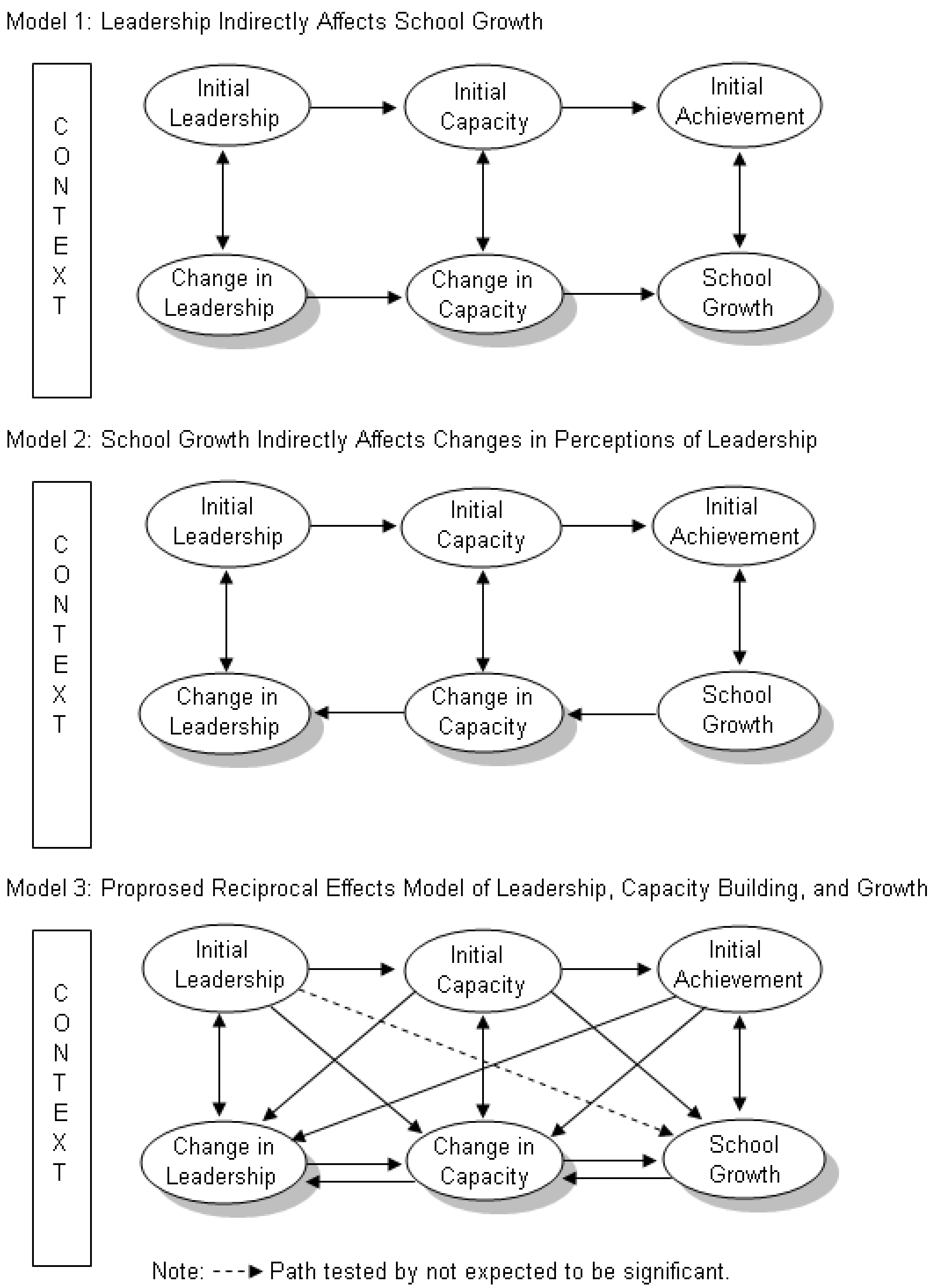
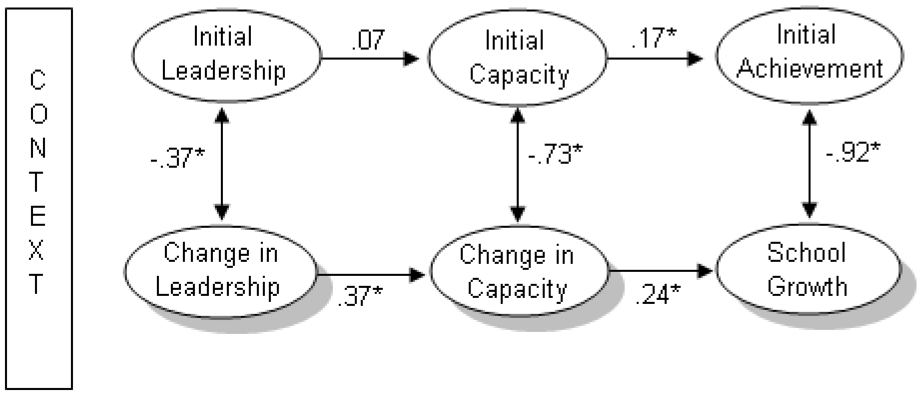
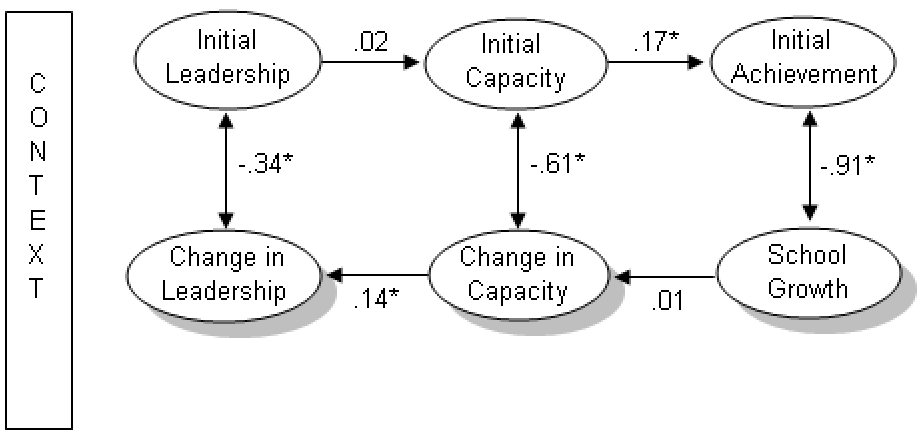


Figure 2
Comparison of Model Estimates

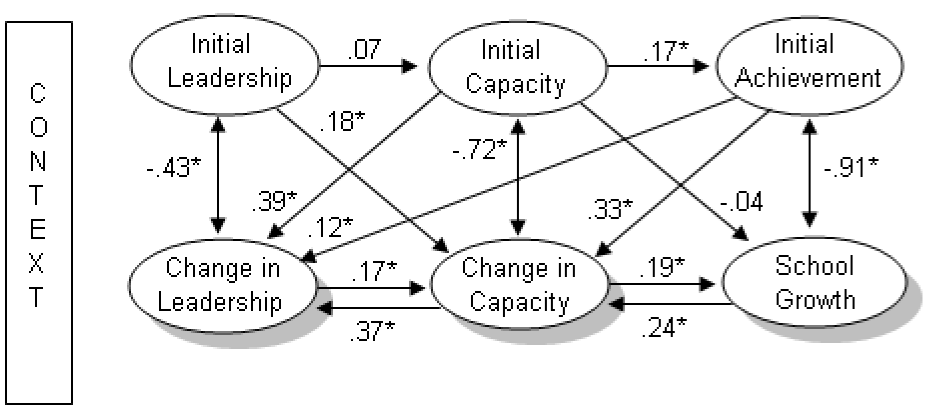
Model 1: Leadership Indirectly Affects School Growth (Note: *$p < .05$)



Model 2: School Growth Indirectly Affects Changes in Perceptions of Leadership



Model 3: Proprosed Reciprocal Effects Model of Leadership, Capacity Building, and Growth



End Notes

[i] In preliminary analyses of these data, we also re-ran our analyses with the parent and student data to extend the generalizability of our basic models.

[ii] For math achievement, the latent change model to represent individual *i* at time *t* can be written as:

$$y_{it} = \nu_t + \Lambda_t \eta_i + K x_i + \varepsilon_{it}, \quad (1)$$

where y_{it} is a vector of math outcomes for individual *i* at time *t* $(y_{i1}, y_{i2}, ..., y_{iT})'$, ν_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* x *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 a covariance matrix (Θ).

After defining the growth portion of the model, the second part of the analysis specifies the relationships between latent variables and other covariates. We can model variability in initial math levels (η_{0i}) and change (η_{1i}) as a function of one or more covariates (x_i) plus error:

$$\eta_{0i} = \alpha_0 + \gamma_0 x_i + \zeta_{0i}, \quad (2)$$

$$\eta_{1i} = \alpha_1 + \gamma_1 x_i + \zeta_{1i}, \quad (3)$$

where α_0 and α_1 are measurement intercepts and γ_0 and γ_1 are structural parameters describing the regressions of latent variables on a covariate (e.g., student SES). Each latent factor has its own residual (ζ_{0i}, ζ_{1i}) that represents errors in the equations for explaining individuals' growth trajectories. The individual student growth model can be further divided into respective individual-level and school-level components (see Muthén and Muthén, 1998-2006). We defined the leadership and school improvement capacity change processes in a similar manner, but only at the school level.

In each proposed model, we grand-mean centered the student-level variables, which results in school-level achievement and growth means which are adjusted for differences in student background. At the school level, we also centered school composition on its respective grand mean for the sample of schools and left principal stability in its raw metric.

[iii] Regression coefficients describing the structural relationships between constructs are contained in the B (Beta) matrix. In this illustration, we will assume the there are three proposed relationships in the B matrix below:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \beta_{13} \\ 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \quad (4)$$

We will assume construct 1 is the dependent variable, in this case, school math achievement. Construct 3 (school improvement capacity) mediates between construct 2 (leadership) and construct 1. There beta matrix specifies that leadership affects school improvement capacity (2 → 3, represented as β_{32}) and, in turn, improvement capacity also affects school achievement (3→1, represented as β_{31}). Moreover, there is a feedback loop present, in that the achievement also affects school improvement capacity (1 →3, represented as β_{31}).

[iv] Because a feedback loop is a causal chain going from one or more latent factor and back to the original factor, one cycle for improvement capacity is $\beta_{13}\beta_{31}$. After two cycles the effect will be $\beta_{13}^2\beta_{31}^2$, and so on. The total effect on school improvement capacity will therefore be the sum of an infinite geometric series, which can be represented as $\beta_{13}\beta_{31}$ /(1- $\beta_{13}\beta_{31}$) for $\beta_{13}\beta_{31}$ < 1 (see Hayduk, 2009 and Jöreskog

& Sörbom, 1989 for further discussion). A sufficient condition for convergence of the series is that the largest eigenvalue of $\mathbf{B}\mathrm{B}'$ is less than one, which can be verified in the model output (Jöreskog and Sörbom, 1989). As Hayduk (2009) observes, however, the assumption of infinite feedback loops versus real-world limitations to the number of cycles that can actually exist may influence the quality of the estimates obtained about the feedback effects. The errors in equations involved in feedback loops (i.e., ζ_1 and ζ_3) are typically specified as correlated (and modeled as as ψ_{13}), which makes sense, since if we assume school improvement capacity and achievement mutually cause each other, then we may expect that they have common omitted causes (Kline, 2004). This type of model corrects the structural parameters between them for possible error in one latent construct that is associated with error in the other (Jöreskog & Sörbom, 1989).

[v] All of the background variables were significant in explaining initial achievement levels and student growth rates (with standardized estimates less than .13), except the relationship between gender and initial student math achievement. Regarding school context controls, student composition was statistically significant in explaining initial school achievement levels, but not growth. Principal stability did not significantly affect school changes in math over time.