

Accepted for Publication in School Effectiveness and School Improvement, 2011

**Conceptual and Methodological Issues in
Studying School Leadership Effects as a Reciprocal Process**

Philip Hallinger¹

Hong Kong Institute of Education

and

Ronald H. Heck

University of Hawaii-Manoa

Abstract

Over the past three decades a substantial body of scholarship has examined the effects of school leadership on student learning. Most of this empirical research has framed leadership as an independent variable, or driver for change, in relation to school effectiveness and school improvement. Yet, scholars have for many years observed that leadership is also influenced by features of the organizational setting in which it is enacted. This leads us to conclude that the predominant approaches to studying school leadership effects provide an incomplete picture of the processes and paths by which leadership contributes to school learning. In this paper, we examine the potential offered by conceptualizations of leadership as a reciprocal, or mutual-influence, process that unfolds over time as an approach to studying leadership for learning. We explore a variety of conceptual and related methodological issues that confront researchers who wish to employ this potentially rich but challenging approach to understanding how school leadership contributes to student learning.

[S]uccessful school principalship is an interactive, reciprocal and evolving process involving many players, which is influenced by, and in turn, influences the context in which it occurs. (Mulford & Silins, 2009, p. 2.)

Mulford and Silins' (2009) characterization of school leadership as a process of reciprocal influence resonates with practitioners and has strong face validity when viewed in light of theoretical treatises on organizational leadership (e.g., Bass & Bass, 2008; Bridges, 1970; Day, Gronn, & Salasc, 2006; Griffin, 1997; Hooijberg & Schneider, 2001; Meindl, 1995; Podsakoff, MacKenzie, & Fetter, 1993; Tate, 2008). Yet, this perspective on leadership is clearly at odds with the preponderance of empirical research on leadership and learning in schools (Hallinger & Heck, 1996b; Heck & Hallinger, 2010). Instead, most scholars have framed leadership, sometimes explicitly but more often implicitly, as an *independent* variable that drives school change and effectiveness (Bossert, Dwyer, Rowan & Lee, 1982; Bridges, 1970, 1977, 1982; Hallinger & Heck, 1996a, 1996b; Leithwood, Day, Sammons, Harris, & Hopkins, 2006). This approach is evident, for example, in frequently cited mediated-effects studies of school leadership effects in the literature (e.g., Hallinger, Bickman, & Davis, 1996; Heck & Hallinger, 2009; Heck, Larson, & Marcoulides, 1990; Krüger, Witziers, & Slegers, 2007; Leithwood & Jantzi, 1999, 2000; Marks & Printy, 2003; Pounder, Ogawa, & Adams, 1995; Wiley, 2001)

Reviews of the school leadership literature conducted over the past 30 years have consistently raised questions about the 'causal ordering' of leadership, school- and classroom-level variables in relation to student learning achievement (e.g., Bossert et al., 1982; Bridges, 1982; Hallinger & Heck, 1996a, 1996b, 1998; Leithwood, Begley, & Cousins, 1990; Leithwood et al., 2006; Pitner, 1988). Causal ordering refers specifically to whether conclusions about 'effective school leadership' reflect valid measurement of leadership impact on learning, or

optimistic attributions derived from correlations or associations between leadership and student achievement (Bridges, 1977, 1982; Hallinger & Heck, 1996b; Pitner, 1988; Rowan, Bossert, & Dwyer, 1983). We note that even the best designed quantitative studies of leadership and learning tend to conclude with caveats related to this issue of causal ordering.

One reason for this limitation is the longstanding and continuing predominance of cross-sectional research in this field of inquiry (Bridges, 1982; Erickson, 1967; Haller, 1979; Hallinger, in press; Hallinger & Heck, 1996a). Unfortunately, cross-sectional research designs are poorly suited to resolving the complex issue of causal ordering of variables in studies that seek to link leadership to learning (Bridges, 1982; Hallinger & Heck, 1996b; Rowan et al., 1983; Witziers, Bosker, & Kruger, 2003). This limitation becomes even more relevant when researchers investigate the impact of leadership on *school improvement*, a process which by definition unfolds over time (Day, Sammons, Leithwood, Hopkins, Harris, Gu, & Brown, 2010; Hallinger & Heck, 1996b; Jackson, 2000; Mulford & Silins, 2003, 2009). Thus, we contend that causal ordering represents an important theoretical issue with practical implications in studies of leadership for learning.

Reciprocal-influence models offer an alternative means of unpacking the issue of causal ordering. They can be traced back to Wright's (1921) formulation of path analytic methods for examining systemic relationships among variables in a proposed theoretical model based on observed correlations. Rather than conceptualizing leaders as the *either* the 'origin or the pawn' (see Bridges, 1970) in leadership for learning, mutual-influence models propose that relationships among variables in a proposed model could have reciprocal or bi-directional effects (Loehlin, 1992). Although reciprocal-influence models have been discussed by scholars for several decades, there have been relatively few empirical tests in either the organizational (e.g.,

Bielby & Hauser, 1977; Blalock, 1970; Duncan, 1969; Duncan, Haller, & Portes, 1968; Goldberger, 1964; Maruyama, 1998; Ployhart, Holtz, & Bliese, 2002; Rogosa, 1980; Sturgis et al., 2004; Williams & Podsakoff, 1989) or educational leadership literatures (e.g., Hallinger & Heck, 2010; Heck & Hallinger, 2010; Mulford & Silins, 2009; Rowan & Denk, 1984).

This paper examines the potential of reciprocal-influence models for illuminating the relationship between school leadership and learning. We first discuss the *conceptual* basis for this approach and then link conceptual models to *methodological* requirements and options. As we shall elaborate, the process of testing reciprocal-influence models through longitudinal data raises a host of new challenges, with relatively few published studies to serve as models (e.g., Hallinger & Heck, 2010; Heck & Hallinger, 2010; Marsh & Craven, 2006; Griffin, 1997; Tate, 2008). With a dearth of useful road maps for exploring this terrain, we felt it imperative not only to identify relevant methodological options, but also to explicate these through the analysis of an illustrative longitudinal data set using structural equation modeling (SEM). We hope that this approach will provide a useful foundation for other scholars who see benefit in exploring the potential of reciprocal-influence models in studies of leadership.

Conceptualizing School Leadership Effects

The term *leadership for learning* implies that causal linkages exist between the intentions and actions of leaders and school learning outcomes. The intellectual lineage of leadership for learning traces back to early attempts to define (e.g., Bridges, 1967; Lipham, 1961) and empirically study (e.g., Gross & Herriot, 1965) models of *instructional leadership* in the USA during the 1960s. As time passed, scholars sought to determine whether school leadership effects on learning could be detected through quantitative research (Bossert et al., 1982; Hallinger & Leithwood, 1994). Most recently researchers efforts have focused on identifying the paths

through which leadership impacts school performance (Hallinger & Heck, 2010; Heck & Hallinger, 2009, 2010; Leithwood, Anderson, Mascall, & Strauss, 2010; Leithwood et al., 2006, 2010; Mulford & Silins, 2009; Robinson, Lloyd, & Rowe, 2008).

Insert Figure 1 about here

This program of school leadership effects research began by employing relatively simple bivariate (i.e., two factor) models that proposed direct effects of school leadership on student learning (see Model A in Figure 1). However, over time, scholars became increasingly critical of the direct-effects approach, asserting that it failed to capture the complexities inherent in the organizational dynamics associated with leadership for learning (Hallinger & Heck, 1996a, 1996b). Researchers proposed more complex models that conceptualized a variety of moderating and mediating variables that could impact this relationship (e.g., Hallinger & Heck, 1996a, 1998, 2010; Leithwood & Jantzi, 1999, 2000; Marks & Printy, 2003; Wiley, 2001). The current state-of-the-art aims at exploring the *paths* through which leadership is linked with student learning (Heck & Hallinger, 2009, 2010; Kruger et al., 2007; Leithwood et al., 2010; Mulford & Silins, 2009). We suggest that mediated- and reciprocal-influence models are foremost among the conceptual models that seek to explore these paths (see Models B and C in Figure 1). Mediating-effect models suggest that a third variable (e.g., school improvement capacity) comes between the independent and dependent variables: that is, the mediator is a function of the independent variable (e.g., leadership) and the dependent variable (e.g., school achievement) is a function of the mediator (MacKinnon, 2008). Reciprocal-influence models build on such key hypothesized relationships between two or three variables as they may unfold over time (Loehlin, 1992; Marsh & Craven, 2006). Thus, this section of the paper is organized around an elaboration of these perspectives on leadership for learning.

Mediated-Effects Models of Leadership for Learning

In recent years, a large body of international research supports the view that school leadership can have significant *indirect* effects on student learning (Bell, Bolam, & Cubillo, 2003; Bossert et al., 1982; Hallinger & Heck, 1998; Leithwood et al., 2006, in press; Leithwood, Patten, & Jantzi, 2010; Robinson et al., 2008; Wiley, 2001). This perspective is represented by Model B in Figure 1. Indeed, scholars have increasingly embraced the belief that school leadership effects on student learning are *mediated* by conditions that build school capacity for change and foster effective teaching and learning (Hallinger & Heck, 1998; Leithwood et al., 2006, in press; Robinson et al., 2008). Empirical evidence, though not conclusive, does provide insight into these paths that link leadership with teaching and learning. Specifically, research indicates that school improvement leadership:

- Impacts conditions that create positive learning environments for students (Hallinger et al., 1996; Hallinger & Heck, 1998; Heck et al., 1990; Heck & Hallinger, 2009; Leithwood et al., 2006, in press; Robinson et al., 2008; Wiley, 2001).
- Mediates academic expectations embedded in curriculum standards, structures, and processes as well as the academic support that students receive (Hallinger et al., 1996; Hallinger & Heck, 1998; Heck et al., 1990; Heck & Hallinger, 2009; Hill & Rowe, 1996; Leithwood & Jantzi, 1999; Leithwood et al., 2006; Robinson et al., 2008).
- Employs improvement strategies that are matched to the changing state of the school over time (Hallinger & Heck, 2010; Heck & Hallinger, 2009, 2010; Leithwood et al., 2006, in press; Mulford & Silins, 2009).
- Supports ongoing professional learning of staff which in turn facilitates efforts of schools to undertake, implement, and sustain change (Fullan, 2006; Hallinger & Heck,

2010; Heck & Hallinger, 2009, 2010; Jackson, 2000; Leithwood et al., 2006, in press; Robinson et al., 2008; Stoll & Fink, 1996).

These descriptions of the *means* by which leadership contributes to school improvement are consistent with what scholars have termed a *mediated-effects* model of leadership (Baron & Kenny, 1986; Pitner, 1988). Leadership effects on learning are produced *indirectly* through their impact on people, structures, and processes in the school that are more proximal to students (Bossert et al., 1982; Hallinger & Heck, 1996a; Leithwood et al., 2006, 2010). While this conceptualization of leadership represents an advance over earlier two-factor studies, we note that mediated-effects models (i.e., Model B in Figure 1) also frame leadership as the *cause* of change in organizational performance. Thus, scholars have noted that mediated-effects models continue to assert, implicitly, a *heroic* role for leaders and fail to take into account the systemic forces and constraints under which they operate (Bossert et al., 1982; Bridges, 1970, 1977, 1982; Meindl, 1995). This assumption was articulated 40 years ago by Bridges, who claimed:

Although administrative man has been described as both the initiator and recipient of action, the dominant focus of the empirical and theoretical work has been on administrative man as an *origin* of his decisions on the one hand, and an *origin* of the behavior of subordinates on the other. . . The understanding we have, in consequence, is limited to the decision making behavior of administrative man as products and processes of a person acting on his own and as a person acting as a causal agent to produce certain effects in the organization. (Bridges, 1970, p. 7)

Moreover, as noted at the outset of this paper, the designs used in mediated-effects leadership studies have been unable to empirically test this assumption of causality. To date,

mediated-effects studies of leadership for learning have, with few exceptions, relied on cross-sectional data. Scholars have, however, noted important limitations when employing cross-sectional research designs to assess the effects of leadership (Bridges, 1982; Haller, 1979; Hallinger & Heck, 1996a, 1996b; Ogawa & Bossert, 1995). For example, in a review of the research conducted 15 years ago we observed that, “[T]he current crop of studies of administrator effects continues to be limited by the persisting reliance on cross-sectional designs. Cross-sectional designs--even ones of high quality--limit our ability to understand the causal relationships involved in studying the impact of school administrators” (Hallinger & Heck, 1996a, p. 36).

More specifically, cross-sectional research designs are neither able to portray adequately the potential interactions that may occur among variables nor determine the direction of causality of proposed relationships. These problems compound when the dependent variable of interest is school change or improvement, both of which imply the need to measure leadership impact over time (Hallinger & Heck, 1996b; Jackson, 2000; Mulford, 2003; Ogawa & Bossert, 1995). Theorists and practitioners know that the nature of life in schools is both complex and cyclical. Understanding how the effects of school processes unfold over a number of school years, therefore, requires research designs and data that are capable of capturing more of this complexity. Recognition of the limitations of cross-sectional snapshots of school processes in school leadership research has led scholars to propose alternative ways of understanding how leadership might impact learning (Hallinger & Heck, 1996a, 1996b, 2010; Heck & Hallinger, 2010; Mulford & Silins, 2003; Pitner, 1988; Southworth, 2002). One alternative model formulation emphasizes reciprocal effects between variable. We summarize this type of mutual causation as Model C in Figure 1 with double arrows between major constructs ($A \rightleftarrows B$).

Reciprocal-Effects Models of Leadership for Learning

As Bridges indicated (1970, 1977), there is a longstanding bias in the leadership literature towards viewing leaders as the source of change in the organization, and either ignoring or understating the extent to which their actions are influenced by the organizational context. We suggest that a more complex set of processes underlies the dynamics of school leadership and school improvement as they unfold over time. These processes extend beyond the efforts of individual leaders seeking to effect change in schools and must take into account how the organization shapes behavior. For this reason, we propose to frame leadership for learning as part of a systemic process that is aimed at impacting student learning (e.g., Fullan, 2006; Ogawa & Bossert, 1995).

The fact that school improvement involves change in the state of the organization over time suggests that the empirical study of leadership for learning requires models that take into account changing relationships among relevant variables. Therefore, as we noted over a decade ago, “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” (Hallinger & Heck, 1996a, p. 19). Indeed, the possibility of reciprocal influence between leaders and followers has been explicitly discussed in the leadership literature for more than two decades (Pitner, 1988; Podsakoff, 1994; Williams & Podsakoff, 1989). Related concepts of reciprocity, responsive adaptation, mutual influence, and leader-follower interaction have been discussed in theories underlying contingency leadership (Fiedler, 1967), servant leadership (Greenleaf, 1977), upward managerial influence (Kipnis et al., 1983), and distributed leadership (Gronn, 2002). Only in the last decade, however, has recognition of mutual influence in the relationship between leaders and followers led to initial empirical tests of reciprocal-effects models in the general leadership

literature (e.g., Griffin, 1997; Keller, 2006; Tate, 2008; Vogelaar & Kuipers, 1997) and in education leadership (Hallinger & Heck, 2010; Heck & Hallinger, 2010; Mulford & Silins, 2009).

However, as Griffin (1997) observed, “The estimation of reciprocal effects has been methodologically problematic in organizational research” (p. 770). A key limiting factor lies in the need for longitudinal data that describe change in organizational processes over time. Such data are, however, difficult to obtain on a scale sufficient to assess the effects of leadership across comparable organizational units. For example, most state and national data bases are not routinely organized in a manner that facilitates longitudinal analyses, though we observe that this is changing with increased policymaker interest in monitoring student growth in learning.

Moreover, researchers have not until relatively recently had access to analytical tools capable of optimally modeling more complex models of mutual influence over time (Heck & Hallinger, in press; Griffin, 1997; Heck & Thomas, 2009; Marsh & Craven, 2006; Podsakoff, 1994; Tate, 2008). This problem is particularly relevant in educational organizations, where studying leadership effects on student learning over time involves dealing with successive measurements that are highly correlated (e.g., achievement data), multiple variables that affect student outcomes, and multiple organizational levels that can affect the environment in which student learning takes place (Hallinger & Heck, 1996b; Hill & Rowe, 1996). Despite these methodological challenges, we cannot overstate the potential advantages of using longitudinal designs in school improvement research where progress has both theoretical implications and practical utility (Griffin, 1997; Heck & Hallinger, in press; Ogawa & Bossert, 1995; Tate, 2008).

Methodological Issues in Examining Reciprocal Influence

There are a number of different ways to model mutual influence utilizing either cross-sectional or longitudinal data (Kline, 2004). In this section of the paper, following Pitner’s (1988)

conceptualization, we explore a ‘cross-lagged’ panel model of reciprocal influence over time. Other mutual-influence models include those with reciprocal (or bidirectional) effects, feedback loops, and parallel growth processes (Kline, 2004; Loehlin, 1992; Muthén & Muthén, 1998-2006). The cross-lagged, autoregressive approach is used in examining two organizational processes over a period of time and represents a potentially powerful means of capturing the effects of leadership on learning and school improvement. We will present different methodological approaches to conducting a cross-lagged, longitudinal analysis of leadership effects on school improvement processes. More specifically, we seek to identify, explore, and illustrate various options for modeling reciprocal relationships between leadership and learning.

Organizational settings are not readily amenable to experimental manipulations. Therefore, research examining temporal relationships is often conducted through longitudinal panel studies in which variables are measured on a number (i.e., at least two) of occasions (Cook & Campbell, 1979). We note, however, that the investigation of a causal process is also a function of the selected research design. Stronger designs (e.g., experiments) are better able to eliminate rival explanations of results and yield more defensible conclusions. Degree of association between proposed variables, relative isolation from other extraneous variables, existence of a temporal relationship (which establishes the direction of influence), and replication of results enhance claims of causal relationships (Bollen, 1989; Campbell & Stanley, 1963; MacKinnon, 2008).

The cross-lagged longitudinal approach to exploring reciprocal effects dates at least from Lazarsfeld and Fiske (1938). They proposed it as an alternative to cross-sectional research for assessing the direction of effects between measured variables in a proposed model. A key

advantage of this approach is that it does not assume instantaneous change in relationships between simultaneously measured variables (Oud, 2002).

Methodology Employed in Testing Illustrative Models

In our effort to elaborate on the conceptual and methodological issues involved in mutual-influence modeling, we refer to a longitudinal data set that describes leadership and learning in schools. The data set consists of survey data collected from teachers in 197 elementary schools on four occasions over a five-year period (i.e., Year 1, Year 3, Year 4, and Year 5). Achievement data were collected from 13,391 third grade students on three occasions (Year 2, Year 3, Year 4).

In years where data from both teachers and students were collected, survey data from schools on collaborative leadership processes and school capacity for improvement were collected *before* student achievement data.ⁱⁱ The temporal sequence of data collection, therefore, makes this data ideal for testing the proposed cross-lagged longitudinal model. Because our interest is in explaining school-level relationships, we focus only on the between-school portion of the models tested in presenting our results. Nonetheless, the proposed models also have within-school models measuring individual students' achievement over a three-year period with a full set of background covariates (e.g., gender, race/ethnicity, socioeconomic status) that was used in estimating school-level achievement at each time point.

SEM offers an ideal methodological framework for investigating complex relationships because of its flexibility in estimating direct, indirect, and reciprocal effects within a single model after accounting for measurement error in measuring the constructs (Heck & Thomas, 2008). In the SEM approach to cross-lagged longitudinal modeling, latent (underlying) variables

are defined at two or more times by a set of observed measures. An important prerequisite is first to establish measurement invariance of the model's latent variables prior to actually testing the longitudinal framework. This implies the constructs being measured have the same meaning over the repeated occasions of measurement (Schlueter et al., 2007). Measurement invariance is generally established by verifying that the same number of latent factors and same factor loadings of observed variables on the factors exist over time. Requiring equal measurement errors of the observed indicators across time is generally considered to be too stringent in establishing measurement invariance (Muthén & Muthén, 1998-2006).

To specify a reciprocal-influence model using SEM, it is important to make sure that we have met model identification rules. These include the following conditions:

- 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 non-redundant elements in the observed covariance matrix;
- Every latent variable with unrestricted variance must emit at least two direct paths to observed indicators or other latent variables when these latter variables have unrestricted error variances (see Bollen & Davis, 2009).

Cross-lagged Modeling of Reciprocal Relationships

Cross-lagged longitudinal models imply that two (or more) variables may be both a cause and an effect of each other over time (Duncan, 1969; Finkel, 1995; Kline, 2004; Marsh & Craven, 2006). Cross-lagged models suggest that the earlier temporal states of component variables (e.g., at Time 1) will mutually reinforce each other over a subsequent interval (or time 'lag'). This reciprocal (or cross-lagged) relationship over two periods of time is defined as $A_1 \rightarrow B_2$ and $B_1 \rightarrow A_2$ (Loehlin, 1992). For example, school leaders may initiate changes in teacher work structures

and expertise, curriculum organization and instructional practices, and student support programs. Changes in the school's educational capacity may produce subsequent effects on leadership behavior, as well as changes in distal outcomes such as student learning. Thus, a reciprocal-influence model can provide evidence to answer questions about whether the proposed relationship between two variables is mutually reinforcing rather than solely unidirectional.

Simple Modeling of Reciprocal Effects with Cross-Sectional Data

As a starting point, we provide a simple illustration of a mutual-influence model using cross-sectional data. Mutual influence implies that two or more variables which are measured at the same time (i.e., within a cross-sectional research design) may be both a cause and effect of each other (Kline, 2004; Marsh & Craven, 2006). If one proposes the existence of a reciprocal effect between two variables, this implies more than a simple cause and effect ($A \rightarrow B$) relationship from the independent variable (A) to the dependent variable (B). It also implies the converse—that B affects A ($B \rightarrow A$). This type of reciprocal effect is indicated with bidirectional arrows ($A \leftrightarrow B$), suggesting that leadership and capacity building mutually affect each other.

Reciprocal relationships cannot be estimated in analyses using multiple regression, since they imply only one dependent variable, but they pose no special difficulties for structural models estimated with iterative model-fitting techniques (Loehlin, 1992). The presence of a reciprocal effect in a structural model makes it non-recursive (Bielby & Hauser, 1977). In terms of diagrammatic representation (as in Model C of Figure 1), it is important to note that reciprocal influence indicated by a bidirectional arrow is not the same as a two-headed arrow ($A \leftrightarrow B$). This latter relationship simply indicates a covariance (or correlation) between the two variables.

In Figure 2, we provide the results of our proposed reciprocal-effects model with a single wave of data from our study. This model fits the data perfectlyⁱⁱⁱ (see Model 1 in Table 1), since

there are just as many model parameters estimated as nonredundant elements in the school-level covariance matrix (Raykov & Marcoulides, 2006). Path coefficients in this figure and the following ones are standardized estimates.

Interpreting reciprocal effects in cross-sectional designs is more difficult than might appear at first glance. Social scientists have long noted that one cannot draw inferences about the direction of causal effects (i.e., causal ordering) from cross-sectional data unless one of the described phenomena clearly preceded the others in terms of their occurrence in time (Kohn, 1977). Our example in Figure 2 illustrates the differences between reciprocal, unidirectional, and correlated relationships implied in a proposed conceptual model. The interpretation of the proposed relationships is that leadership influences school improvement capacity (0.46, $p < .05$). In turn, improvement capacity influences outcomes (0.35, $p < .05$), which simultaneously is influenced by the state of the school's academic outcomes (0.23). In this case, however, the latter relationship was not statistically significant ($p > .05$). The non-significant result for this latter path is due to a large standard error in measuring the between-school effect of student achievement on improvement capacity. This implies considerable variability in the size of the effect across the sample of schools.

As shown in the figure, the errors in equations involved in a reciprocal relationship (i.e., represented as short single-headed arrows) are typically specified as correlated (although we had to restrict this path to 0.0 in this simple example in order to achieve model identification). Correlated residual variances is consistent with the logic of reciprocity, since if we assume that A and B mutually influence each other, we may reasonably expect that they have common omitted causes (Anderson & Williams, 1992; Kline, 2004; Loehlin, 1992). We could, of course, extend the logic of reciprocal relationships by suggesting a type of feedback loop in the model, where

leadership affects improvement capacity, improvement capacity affects achievement outcomes, and the state of the school's achievement might influence school leadership actions.

Insert Figure 2 about Here

In cases where one can realistically assume the presence of mutual causation, our model demonstrates that it is possible to assess the magnitude of these reciprocal effects using SEM techniques for solving simultaneous equations. However, it should be emphasized that analyses of reciprocal relationships with cross-sectional data can only display a portion of the more complex relationship that may exist between the variables as they interact *over time*. Testing reciprocity in relationships with cross-sectional data is subject to severe limitations, since the operationalized model lacks relevant information about the *temporal relationships* among the variables.

Assessment of reciprocal causation among a set of variables measured at the same time requires an assumption of equilibrium. More specifically, one must assume that the relationship specified in a reciprocal-relationship between variables A and B (or among variables in a feedback loop) has already manifested its effects, and therefore, the system is essentially in a balanced state (Kline, 2004). This means that its estimation does not depend on the particular time in which the data were collected (Kline, 2004). Violation of the equilibrium assumption can lead to biased estimates, and in the world of organizations we suggest that this assumption of stability is often difficult to justify. A stronger approach, therefore, is to explore mutual influence between variables through the use of longitudinal data.

Exploring Reciprocal Relationships with Two or More Waves of Data

Although it is possible to test reciprocity in relationships using cross-sectional data, as we have noted the limitations are considerable. One primary limitation concerns the requirement that

the system is stable when it is observed, such that the reciprocal effects observed between two variables will be similar at whichever point in time that they are observed. Yet, the validity of this assumption can never be ascertained with a single round of data collection. We suggest, therefore, that reciprocal interaction entails a clear assumption that behavioral response and adaptation unfold over time (Griffin, 1997; Hallinger & Heck, 1996b; Heck & Hallinger, in press; Mulford & Silins, 2009; Tate, 2008).

Ogawa and Bossert (1995) summarize the case for using longitudinal data in studies of leadership effects:

[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)

With this in mind, we suggest that longitudinal data are an imperative if we seek to define and test organizational models that propose reciprocal effects. Therefore, a more robust way to specify reciprocal effects lies in the use of longitudinal panel designs in which variables are each measured on two or more different occasions (Bollen & Curren, 2006; Lazarsfeld & Fiske, 1938; Marsh & Craven, 2006; Williams & Podsakoff, 1989). More specifically, we suggest that longitudinal modeling of organizational processes offer clear advantages over the models with feedback loops obtained using cross-sectional data. These include a temporal ordering of latent variables related to measurement occasions and the ability to measure stability versus change in

levels of each latent variable as well as variability in the mutually-reinforcing effects over time (Kline, 2004). However, these longitudinal models cannot be analyzed optimally by ordinary least squares (OLS) regression due to correlated errors between observations.

Specifying a cross-lagged longitudinal model. As we have emphasized, longitudinal models are better able to test proposed mutually-reinforcing relationships. We first develop an autoregressive cross-lagged model to investigate the possibility that leadership and school improvement capacity are mutually-reinforcing constructs, each leading to gains in the other. As summarized in Figure 3, reciprocal causation can be represented by cross-lagged direct effects between A and B measured at different times. For ease of presentation, we only focus on relationships between the latent constructs and not the measurement model (i.e., which consists of the observed indicators of each latent variable and their corresponding residual errors).^{iv}

The model is based on the assumption that each latent construct η in organization i measured at time t is a function of its lagged value at time $t-1$ (the autoregressive effect), plus the lagged value of another latent construct measured at time $t-1$ (the cross-lagged effect), plus error (Finkel, 1995; Schluetler et al., 2007). More specifically, an earlier state of A (which we will label η_{Lit-1} in the figure to indicate the measurement of the latent leadership construct at Time 1 or T_1) affects the subsequent state of B (which we will label η_{Cit} to indicate the measurement of the latent improvement capacity construct at Time 2 or T_2) and, simultaneously, an earlier state of improvement capacity (η_{Cit-1} at T_1) affects the later state of leadership (η_{Lit} at T_2). The autoregressive effect of each latent variable allows an assessment of the stability of each construct over time—that is, the changes in the rank order of individuals between the two points in time, as opposed to absolute changes in their scores (Schlueter et al., 2007). Note, however,

that the cross-lagged formulation does not directly test whether the two variables are causally related within each simultaneous period of data collection. However, if one of the variables under consideration were measured before the other within each period of data collection, we could add this relationship to the proposed model.

Insert Figure 3 about Here

An initial (T_1) covariance (or correlation) between the constructs is also proposed in Figure 3. There is also a covariance proposed (curved two-headed arrow) between the residual variances of the constructs within the same measurement occasion (i.e., represented by short arrows for each construct). The residuals associated with latent variables represent errors in predicting their status at time t from time $t-1$. The residual for each observed indicator of the latent variable is allowed to covary with itself across the measurement occasions. These relationships are referred to as autocorrelated measurement errors. Failure to account for covariation between errors can bias the model estimates of mutual influence (Sturgis et al., 2004; Williams & Podsakoff, 1989). Because they are part of the measurement model, they are not shown in Figure 3 or in subsequent figures.

For two latent variables (leadership and improvement capacity) at two points in time, the structural relationships for the latent-variable autoregressive cross-lagged model for individual (or in this case school) i at time t can be defined as follows:

$$\begin{aligned}\eta_{Lit} &= \alpha_{Li} + \beta_{L1}\eta_{Lit-1} + \beta_{C2}\eta_{Cit-1} + \zeta_{Lit} \\ \eta_{Cit} &= \alpha_{Ci} + \beta_{C1}\eta_{Cit-1} + \beta_{L2}\eta_{Lit-1} + \zeta_{Cit},\end{aligned}\tag{1}$$

where α_{Li} and α_{Ci} are intercepts, η_{Lit-1} and η_{Cit-1} represent the latent leadership and improvement capacity constructs for individual i at time $t-1$, β_{L1} and β_{C1} are autoregressive structural coefficients, β_{L2} and β_{C2} are cross-lagged structural coefficients, and ζ_{Lit} and ζ_{Cit} are errors in predicting each outcome at Time 2.

The magnitude of the cross-lagged coefficients indicates how much variation in η_{it-1} predicts aggregate changes in η_{it} , controlling for autoregression of each latent construct (Schlueter et al., 2007). The standardized cross-lagged effects can then be compared. These relationships are proposed to be similarly related at T_3 (and subsequent occasions). We show this formulation over three waves of data in Figure 4, which provides the results of this model test.

The cross-lagged longitudinal approach is not without criticism (e.g., Oud, 2002; Rogosa, 1980). A key limitation is that in discrete-time models different time lags can result in different estimates of effects. We did find this occurs in our example models. If time intervals are poorly chosen in cross-lagged models, it is possible to miss the effect because the interval chosen was either too short or too long (so the effect faded). Since no one lag is sufficient to fully reveal the nature of a proposed causal relationship, the use of varying lags within or across studies should be considered (Selig & Preacher, 2009). Although real-life events do unfold continuously over time, the availability of relevant data is often a challenge since they may only be collected at particular intervals (e.g., such as yearly). We applied suggestions for varying the length of causal lags, as well as testing for invariance of structural parameters across the time period of the study. This yields an average effect in the observed relationships across the period of time under consideration (Sturgis et al., 2004). An alternative approach to addressing to this problem is through continuous-time modeling (e.g., see Oud, 2002).

A second limitation is that the cross-lagged approach is also not as adaptable to flexible treatments of time as, for example, latent change models. The latter are able to provide individually-varying trajectories across organizational settings and incorporate nonlinear growth (Sturgis et al., 2004). We could also represent a cross-lagged model with at least three measurement occasions as a *latent curve* model. This type of model has some advantages where the aim of the analysis is to describe each school's change trajectory over time. For example, we could distinguish schools that underwent considerable growth in capacity building or achievement from others whose growth patterns in these domains was flat or in decline. In this case, however, we were primarily interested in describing the stability of patterns of mutual influence across several iterations of data collection, so the cross-lagged longitudinal model seemed well suited to this purpose. It was also capable of displaying possible direct and indirect effects over time.

For cross-lagged models, the means represent the average at each measurement occasion. For example, the leadership intercepts (in this case, percentage of agreement) for Times 1-3 were 0.76, 0.76, and 0.69 respectively (not tabled). For capacity, the factor score intercepts were 0.0, 0.05, and 0.21 for Times 1-3, respectively (not tabled). The average school math achievement scores were 217.5, 235.4, and 250.7, respectively. Of course, different schools could have quite different individual trajectories on each of the three constructs.

Establishing measurement invariance. As we noted earlier, it is important to ensure that the measurement model (i.e., consisting of the items that define each construct on each occasion) displays measurement invariance. We tested the measurement model (consisting of the latent factors, their observed indicators, and errors) initially and found that it was invariant (i.e., invariant factors and factor loadings) over the three occasions of measurement (see Table 1).^v

We also tested the hypothesis of equal errors for observed items across time (using the change in chi-square coefficient for the associated degrees of freedom between the two models; we rejected this hypothesis because the change in chi-square coefficient was large ($\Delta\chi^2$, 8 df, = 62.36, $p < .05$, not tabled).

We also note in passing that we could also represent cross-lagged effects with at least three measurement occasions as a *latent change* model (Heck & Hallinger, 2009, in press). This latter type of model has some advantages where the aim of the analysis is to describe each school's change trajectory over time—for example, schools that underwent considerable growth in capacity building or achievement versus others whose growth in these domains was more flat or declined.

Testing a mutually-reinforcing relationship. When we test Model 2 with three waves of school data, we see in Table 1 that it provides an adequate fit to the data using selected fit indices (as summarized in endnote iii). For example, the CFI is 0.96, slightly above an often-accepted standard of 0.95 (Raykov & Marcoulides, 2006). Figure 4 summarizes the standardized structural relationships, as well as the invariant factor loadings across the three measurement occasions for each of the four indicators of school improvement capacity. As the figure shows, the errors for each indicator defining improvement capacity are not invariant over time. Because there is only one observed scale defining collaborative leadership (consisting of eight items), its factor loading is invariant by definition, since one unstandardized factor loading must be fixed at 1.0 for each latent variable to define a metric measuring for the factor (and its error term is also fixed at 0.0 across occasions).

Insert Figure 4 about Here

The cross-lagged relationship proposed between leadership at T₁ and improvement capacity at T₂ is significant ($\beta = .10, p < .05$), as is the relationship between improvement capacity at T₁ and leadership at T₂ ($\beta = .26, p < .05$). The cross-lagged relationship for leadership \rightarrow capacity ($\beta = 0.10, p < .05$) is the same between T₂ and T₃. The relationship from capacity \rightarrow leadership is considerably stronger between T₂ and T₃ ($\beta = 0.72, p < .05$) than between T₁ and T₂. Because the length of time is less between T₂ and T₃ (one year) than between T₁ and T₂ (2 years), this may be evidence that the relationship strengthens over time, but we cannot answer this definitely in this model. As Figure 4 indicates, there does not seem to be a cross-lagged relationship between T₁ and T₃ (i.e., which represents four years between school surveys). We note the presence of significant relationships between two variables increases the credibility of a causal relationship but is not sufficient to establish the necessary temporal antecedence-consequence relationship (Cook & Campbell, 1979).

The stability relationships between leadership at T₁ and T₂ ($\beta = 0.51$) and between capacity at T₁ and T₂ ($\beta = 0.76$) are also significant ($p < .05$). Similarly, between T₂ and T₃ they are also significant, although the stability of leadership from T₂ to T₃ is considerably weaker ($\beta = .14, p < .05$). This weaker stability coefficient corresponds with the fact that leadership perceptions on average dipped few percentage points between T₂ and T₃. We note also that improvement capacity at T₁ also exerts a weak, but positive, effect on improvement capacity at T₃ ($\beta = .08, p < .05$). The correlation between leadership and academic capacity at T₁ is 0.18, and the correlation between the residuals at T₂ is 0.40 (at T₃, the correlation is stronger at 0.91). An alternative explanation could also be that the results observed may also be related to principal stability. We subsequently tested the plausibility of this, by adding principal stability as a control

on leadership and capacity building over time in the full model but found it did not affect the stability of the results.

We also tested for the equality of the stability coefficients and cross-lagged coefficients across measurement occasions but found these hypotheses were not supported, since the change in chi-square coefficient for the associated degrees of freedom was large in each case (not tabled). Variation in the size of the cross-lagged relationships over time provides initial empirical support for the view that leadership and improvement capacity are mutually reinforcing over time. We can also see that previous improvement capacity appears more strongly related to subsequent leadership at each interval, than the relationship between previous leadership and subsequent improvement capacity. Although unequal cross-lagged coefficients support that the relationship between leadership and improvement capacity is not spurious over time, it would be a mistake to argue that their relative magnitude indicates the causal predominance of one variable over the other (Rogosa, 1980). In Figure 4 we note that the residual variances for leadership and improvement capacity (i.e., curved two-headed arrows) were correlated at T_2 and T_3 .

Proposing the full model. We are not aware of any educational leadership models that have been examined empirically as suggested in the previous figures. In Model 3 we add math achievement (see Figure 5). The figure conceptualizes leadership as both an independent and dependent variable embedded within the organizational context. More specifically, leadership is still viewed as impacting school improvement in learning; however, the relationship is viewed as primarily through *indirect* paths (Hallinger & Heck, 1996b). Such mediated-effects models examine whether a hypothesized cause-effect relationship can be better explained by specifying a construct that is more closely related to the outcome (Calsyn, Winter & Burger, 2005). We show the hypothesized effect of leadership on subsequent achievement with *dotted* lines in Figure 5 to

emphasize that we expect the paths to be non-significant, but we test for the presence of possible *direct* effects of previous leadership at each occasion. For ease of presentation, once again we do not show the observed indicators and error terms for each measurement occasion.

Insert Figure 5 about Here

The key difference between Model 3, as shown in Figure 5, and the mediated-effect model in Figure 1 (i.e., Model B) lies in the modeling of subsequent leadership behaviors over time in *response* to other organizational variables and changes in student learning and then measuring their *subsequent effect* on these other school processes later in the study. In our proposed model, school improvement capacity was also proposed as a critical mediating effect between organizational leadership and school outcomes. We could then speculate that the leadership effects on school-level processes would again affect outcomes levels in the future.

This type of model offers a potentially important advantage in leadership research where we are interested in exploring how leadership may adapt under different situations, or contingencies, that may develop over time (Hallinger & Heck, 2010; Heck & Hallinger, 2009, in press; Ogawa & Bossert, 1995). The model in Figure 5 also implies that time-invariant *contextual* variables likely interact with the organizational processes under consideration (which are incorporated into the model tests), but we do not hypothesize specific relationships. The school controls included student composition, number of students enrolled, teaching staff experience, professional qualifications, staff stability, and principal stability.

Testing the full model. Once again, when we test the model (see Model 3 in Table 1), we find that it also provides an adequate fit to the data. In this case, the CFI is 0.95, consistent with our accepted standard of 0.95 (e.g., Raykov & Marcoulides, 2006). The standardized path

coefficients are summarized in Figure 6. For ease of presentation, again we do not include the observed variables and error terms for each measurement occasion in the between-schools model (which are the same as those in Figure 4).

Insert Figure 6 about Here

We first note that the mutually-reinforcing relationship between leadership and capacity building is again supported over the three measurement occasions. Second, although tested, the figure suggests that prior leadership is not directly related to subsequent achievement at each measurement occasion. In contrast, prior improvement capacity is related to subsequent achievement at each occasion (with standardized coefficients ranging from 0.23 at T₁ to 0.08 and 0.09 at T₂ and T₃, respectively ($p < .05$)). We also note that improvement capacity at T₂ was related to subsequent math achievement at T₃ ($\beta = .13, p < .05$), but improvement capacity at T₁ was not related to achievement at T₂.

Figure 6 also indicates that prior achievement at each occasion was positively related to subsequent levels of improvement capacity (with $\beta = 0.12$ at T₂ and $\beta = 0.19$ at T₃). Moreover, math achievement at T₁ was not predictive of leadership at T₂ ($p > .05$); however, math achievement at T₂ was predictive of leadership at T₃ ($\beta = 0.23, p < .05$). Once again, we note that the residual variances for leadership and improvement capacity (i.e., curved two-headed arrows) were correlated at Time 2 and Time 3. We also tested whether the residual variances for math at Time 2 and Time 3 were correlated with the residual variances for leadership or improvement capacity at Time 2 and Time 3 but found there was no association.

Finally, only a limited number of covariates in the model (i.e., student composition, staff stability, teacher professional qualifications) exerted small effects (i.e., standardized coefficients

$< .25, p < .05$) on the constructs on one or more measurement occasions. Including them, however, improves the examination of the model's component constructs.

Extending the Reciprocal-Effects Model

Empirical support in Figure 6 for the mutual-influence model supports the theoretical premise that changes in organizational leadership and school improvement capacity represent a mutually-reinforcing organizational process. This suggests the process is one in which the organization “gains momentum” over time through changes in leadership and academic capacity that are organic and mutually responsive. We can extend the validity of the proposed model in Figures 2 to 6 by examining whether it is useful in explaining variation in the leadership and capacity constructs at a fourth point in time. To conduct this test, we drew upon a subsequent round of survey data that became available after our primary model was developed and tested.

When we examine this final model (Model 4 in Table 1), we find that it continues to be consistent with the observed data ($CFI = 0.95$). In Figure 7, the standardized coefficients indicate that leadership at T_3 is related to improvement capacity at T_4 ($\beta = 0.09, p < .05$) and vice versus ($\beta = 0.16, p < .05$). The stability coefficients for each construct are also strong between T_3 and T_4 ($\beta > 0.60, p < .05$). This suggests that leadership was more stable between T_3 and T_4 than between T_2 and T_3 . One explanation is that the average leadership intercept is similar between T_3 and T_4 (i.e., rising slightly from 0.69 to 0.71).

Insert Figure 7 about Here

Overall, the model coefficients at T_4 offer further support for the premise that leadership and school improvement capacity represent a mutually-reinforcing organizational system—and that the system seems in relative stability over time. We also find that achievement at T_3 is

positively related to subsequent changes at T₄ in *both* leadership and capacity building. In this final model, we also note some preliminary support for the hypothesis of indirect effects of previous leadership on subsequent math achievement. More specifically, there is evidence of a small *indirect* effect of previous leadership on math at T₂ through combined paths ($p < .05$), and there is some support for this same hypothesis at T₃ also ($p < .08$). Once again, we estimated this model controlling for principal stability. Finally, we note that the final model accounted for considerable variance (i.e., roughly 50-60%) in the component constructs at T₃ and T₄.^{vi}

Conclusions

This paper is located within the intellectual lineage of research that studies school leadership effects (Bossert et al., 1982; Gross & Herriot, 1965; Hallinger & Heck, 1998; Pitner, 1988; Robinson et al., 2008). Scholars claim that significant progress has been made in understanding the nature of leadership effects on school improvement and student learning.

Leithwood and his colleagues recently summed up this position:

School leaders are capable of having significant positive effects on student learning and other important outcomes. . . . Indeed, enough evidence is now at hand to justify claims about significant leadership effects on students that the focus of attention for many leadership researchers has moved on to include questions about how those effects occur. (Leithwood et al., 2010, p.1)

The body of empirical research on leadership for learning has progressed from the use of relatively simple towards more complex conceptual and analytical models (Bridges, 1967; Lipham, 1961; Hallinger & Heck 1996a, 2010; Leithwood et al., 2010; Marks & Printy, 2003). In an earlier review of research on school leadership effects, we explicitly advised scholars to

forego direct-effects studies of leadership and learning (Model A in Figure 1) in favor of mediated- and reciprocal-effects models (Models B and C in Figure 1; Hallinger & Heck, 1996a, 1996b). Subsequent reviews of this research suggest that researchers have largely moved towards the adoption of mediated effects models in doctoral research (Hallinger, in press) as well as in the published literature (Leithwood et al., 2006; Robinson et al., 2008). Indeed, we wish to suggest that the research community has settled in, perhaps too comfortably, with conceptualizations and techniques involved in the use of mediated-effects models. In support of this assertion, we noted earlier in this paper that there have been no more than a handful of empirical studies that have employed reciprocal-influence models to the study of leadership and school improvement (e.g., Hallinger & Heck, 2010; Heck & Hallinger, in press; Mulford & Silins, 2009; Rowan & Denk, 1984).

This paper has sought to review the conceptual basis for employing reciprocal-influence models of leadership and learning, and demonstrate alternative methodological approaches for use in empirical studies. The illustrations provided in this paper suggest that reciprocal-effects modeling does have potential to reveal additional information about the nature of relationships among relevant variables in models of leadership for learning. This information is essentially ignored when we rely upon widely accepted, unidirectional, mediated-effects analyses (see also Marsh & Craven, 2006; Hallinger & Heck, 2010; Heck & Hallinger, in press). Thus, we assert that the analysis of longitudinal data within a reciprocal-effects framework may provide a complementary and, perhaps, more comprehensive picture of the processes at work in leadership for learning. Finally, we also suggest that the empirical results which derive from this analytical approach are grounded in a potentially richer theoretical and more practical perspective on the dynamic role of leadership in school improvement. Rather than framing leadership as a ‘heroic’

agent of change, this perspective offers a path towards the study of leadership as both a cause and effect of school improvement processes.

We note that our results should be considered along with several limitations. Because of their increased complexity, cross-lagged models tested are more difficult to estimate. Consequently, it becomes essential to examine proposed solutions for illogical parameters (e.g., negative error variances, large standard errors). Consistent with others' use of this approach, we found that differences in the choice of discrete time intervals can also influence the pattern of results. This can make it more difficult to determine whether the overall system being studied is stable over time (i.e., by testing hypotheses of equal stability coefficients and cross-lagged coefficients over time). In our illustration, although we confirmed measurement invariance of the factors and factor loadings over time, we could not make the same claim for the structural coefficients in the model. Because this particular approach to reciprocal-influence modeling focuses on average effects between time intervals, it ignores the variability in individual schools' pattern of change over time. When the researcher's interest lies in examining the growth trajectories of individual schools in more detail (i.e., as in comparing different patterns of growth or change), latent change modeling is a more flexible modeling approach (Heck & Halliner, 2009; Heck & Thomas, 2009). Developing separate trajectories for each unit potentially allows the analyst to uncover more varied patterns of change from the average, or typical, trajectory observed in the sample. We refer the reader to other reciprocal-effects analyses that employed latent change modeling (Heck & Thomas, 2009).

It is also important to acknowledge that reciprocal-influence models may still not resolve issues of whether variable A causes B or variable B causes A, unless relevant limitations are minimized. As noted, another relevant condition is that relationships among variables within the

causal structure are relatively stable. By examining our reciprocal-effects model over four time periods, we were able to provide initial support for the *assumption* that the structure of the proposed relationships was relatively stable over time. In contrast, the stability of the organizational system can only be assumed but not verified in cross-sectional tests of the model. It is still important to note, however, that even this type of test does not provide complete protection against rival explanations, for example, a possible selection-bias argument (Cook & Campbell, 1979). For example, teachers may perceive improvement capacity more positively in schools that with high achievement over long periods of time.

In addition, caution must be exercised in using SEM applications to test substantive theories. Omission of relevant variables often produces misleading interpretation of results (Kline, 2004). Although we included a fairly comprehensive range of school context variables in our model investigations, it is likely that other educational processes would also influence student learning. These might include teacher classroom behavior, student grouping strategies, and student academic and social integration within the school. For these reasons, it is best to consider results from non-experimental, cross-lagged panel designs as preliminary rather than definitive (Campbell & Stanley, 1966).

Despite these limitations, we are impressed with the potential that reciprocal-effects models offer for the study of leadership, learning and school improvement. Indeed, we believe the application of reciprocal-influence models provides a useful complement to the extant set of unidirectional, mediated-effects studies of school leadership effects. We encourage other researchers working in this area of education research to explore the potential of these varied models as a means of clarifying and expanding our understanding of the relationship between leadership and learning.

References

- Anderson, S. E. & Williams, L. J. (1992). Assumptions about unmeasured variables with studies of reciprocal relationships. The case of employee attitudes. *Journal of Applied Psychology, 77*, 638-650.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.
- Bass, B. (2008). *Bass and Stogdill's handbook of leadership*, 3rd edition. New York: Free Press.
- 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.
- Berry, W. D. (1984). *Nonrecursive causal models*. Thousand Oaks, CA: Sage.
- Bielby, W. T. & Hauser, R. M. 1977. Structural equation models. *Annual Review of Sociology, 3*, 137-161.
- Blalock, H. M. Jr. 1970. Estimating measurement error using multiple indicators and several points in time. *American Journal of Sociology, 35*, 101-111.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. & Davis, W. R. (2009). Two rules of identification for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 523-36.
- 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. (1967). Instructional leadership: A concept re-examined. *Journal of Educational Administration, 5*(2), 136-147.
- Bridges, E. (1982). Research on the school administrator: The state-of-the-art, 1967-1980. *Educational Administration Quarterly, 18*(3), 12-33.

- Bridges, E. (1977). The nature of leadership. In L. Cunningham, W. Hack, and R. Nystrand (Eds.), *Educational administration: The developing decades* (pp. 202-230). Berkeley: McCutchan.
- Calsyn, R. J., Winter, J. P., & Burger, G. K. (2005). The relationship between social anxiety and social support in adolescents: a test of competing causal models. *Adolescence*, 40(157), 103-13.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin company.
- Cook, T. & Campbell, D. T. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Boston: Houghton Mifflin Co.
- Day, D., Gronn, P., & Salasc, S. (2006). Leadership in team-based organizations: On the threshold of a new era. *The Leadership Quarterly*, 17(3), 211-216.
- Day, C., Sammons, P., Leithwood, K., Hopkins, D., Harris, A., Gu, Q., & Brown, E. (2010). *Ten strong claims about successful school leadership*. Nottingham, UK: The National College for School Leadership.
- Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, 72, 177-182.
- Duncan, O. D., Haller, A. O., & Portes, A. 1968. Peer influences on aspirations: A reinterpretation. *American Journal of Sociology*, 74, 119-137.
- Erickson, D. (1967). The school administrator. *Review of Educational Research*, 37(4), 417-432.
- Fiedler, F. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.
- Finkel, S. (1995). *Causal analysis with panel data*. London: Sage.
- Fullan, M. (2006). *Turnaround leadership*. New York: Wiley & Sons.
- Goldberger, A. S. 1964. *Econometric theory*. New York: Wiley.
- Greenleaf, R. (1977). *Servant leadership*. Mahwah, N.J.: Paulist press.
- 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.
- Gross, N., & Herriot, R. (1965). *Staff leadership in schools*. New York: Wiley.
- Haller, E. (1979). Questionnaires and the dissertation in educational administration. *Educational Administration Quarterly*, 15(1), 47-66.
- Hallinger, P. (In press). A review of three decades of doctoral studies using the *Principal Instructional Management Rating Scale*: A lens on methodological progress in educational leadership. *Educational Administration Quarterly*.
- 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. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership and Management*, 30(2), 95-110.
- Hallinger, P., & Heck, R.H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157-191.
- Hallinger, P., & Heck, R.H. (1996a). 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.H. (1996b). The principal's role in school effectiveness: An assessment of methodological progress, 1980-1995. In K. Leithwood et al. (Eds.), *International handbook of research in educational leadership and administration* (pp. 723-784). New York: Kluwer Press.
- Hallinger, P. & Leithwood, K. (1994). Exploring the impact of principal leadership. *School Effectiveness and School Improvement*, 5(3), 206-218.
- 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). Collaborative leadership effects on school improvement: Integrating unidirectional- and reciprocal-effects models. *The Elementary School Journal* 111(2), 226-252.
- Heck, R.H., Larson, T., & Marcoulides, G. (1990). Instructional leadership and student achievement: Validation of a causal model. *Educational Administration Quarterly*, 26(2), 94-125.
- Heck, R. H., & Thomas, S. (2009). *An introduction to multilevel modeling* (2nd edition). New York: Routledge.
- Heise, D. R. (1975). *Causal analysis*. New York: John Wiley & Sons.
- Hill, P., & Rowe, K. (1996). Multilevel modeling in school effectiveness research. *School Effectiveness and School Improvement*, 7, 1-34.
- Hooijberg, R., & Schnieder, M. (2001). Behavioral complexity and social intelligence: How executive leaders use stakeholders to form a systems perspective. In S. Zaccaro and R. Klimoski (Eds.), *The nature of organizational leadership: Understanding the performance imperatives confronting today's leaders* (pp. 104-131). San Francisco: Jossey-Bass.
- Jackson, D. (2000). The school improvement journey: Perspectives on leadership. *School Leadership & Management*, 20(1), 61-78.
- 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.
- Kohn, M. L. (1977). *Class and conformity: A study in values, with a reassessment*. Chicago: University of Chicago 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.
- Lazarsfeld, P. F. & Fiske, M. (1938). The panel as a new tool for measuring opinion. *Public Opinion Quarterly*, 2, 596-612.
- Leithwood, K., Anderson, S., Mascall, B., & Strauss, T. (2010). 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.
- Leithwood, K., Begley, P. & Cousins, B. (1990). The nature, causes and consequences of principals' practices: A agenda for future research. *Journal of Educational Administration*, 28(4), 5-31.
- 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. & 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. & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with the school. *Journal of Educational Administration*, 38(2),112-129.
- Leithwood, K., Patten, S., & Jantzi, D. (2010). *Testing a conception of how school leadership influences learning*. Unpublished paper. Toronto, CA: U. of Toronto/OISE.
- Lipham, J. (1961). *Effective principal, effective school*. Reston, VA: National Association of Secondary School Principals.
- Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis* (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Taylor & Francis.

- 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.
- Maruyama, G. M. (1998). *Basics of structural equation modeling*. Newbury Park, CA: Sage.
- Mathieu, J., Ahearne, M., & Taylor, S. (2007). A longitudinal cross-level model of leader and salesperson influences on sales force technology use and performance. *Journal of Applied Psychology, 92*(2), 528-537.
- Meindl, J. (1995). The romance of leadership as follower-centric theory: A social constructionist approach. *The Leadership Quarterly, 6*, 329-341.
- 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.
- Oud, J. H. L. (2002). Continuous time modeling of the cross-lagged panel design. *Kwantitatieve Methoden, 69*, 1-26.
- Pitner, N. (1988). The study of administrator effects and effectiveness. In N. Boyan (Ed.), *Handbook of research in educational administration* (pp. 106-132). 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.

- 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.
- Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, 88(2), 245-258.
- Rowan, B., Bossert, S., & Dwyer, D. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, 12, 24-31.
- Rowan, B., & Denk, C. (1984). Management succession, school socioeconomic context and basic skills achievement. *American Educational Research Journal*, 21(3), 17-537.
- Schlueter, E., Davidov, E. & Schmidt, P (2007). Applying autoregressive cross-lagged and latent growth curve models to a three-wave panel study. In K. Montfort, J. Oud, and A. Satorra (Eds.) *Longitudinal models in the behavioral and related sciences* (pp. 315-337). New York: Routledge.
- Selig, J. P. & Preacher, K. P. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6(2-3), 144-164.
- Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership and Management*, 22(1), 73-92.
- Stoll, L., & Fink, D. (1996). *Changing our schools: Linking school effectiveness and school improvement*. Buckingham: Open University Press.
- Sturgis, P., Smith, P., Berrington, A., & Hu, Y. (2004). *Using repeated measures data to analyse reciprocal effects: The case of economic perceptions and economic values*. Conference Papers from the Department of Sociology, Paper 4. Downloaded from <http://epubs.surrey.ac.uk/publsoc/4>
- 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.

- Teece, D. (1982). Towards an economic theory of the multiproduct firm. *Journal of Economic Behavior and Organization*, 3, 39-63.
- 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.
- 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.
- Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39, 398-425.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.

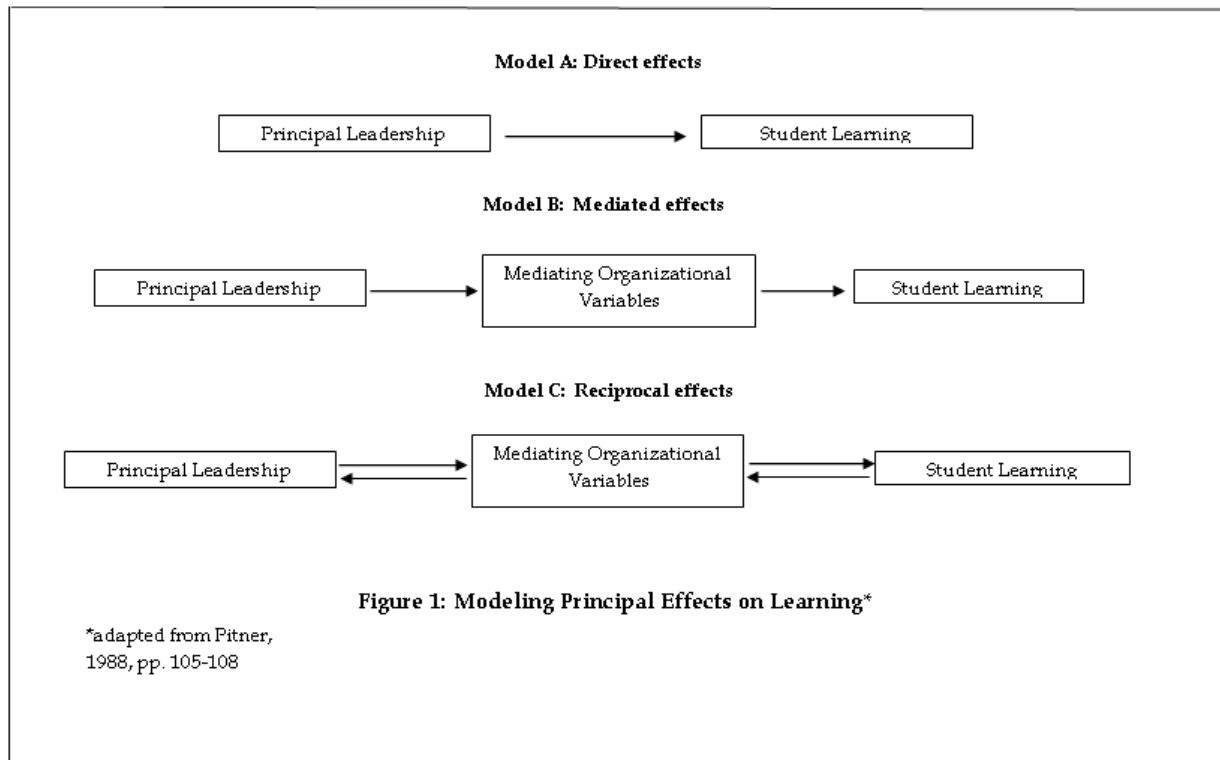


Figure 1. Conceptual Models of Leadership and Learning

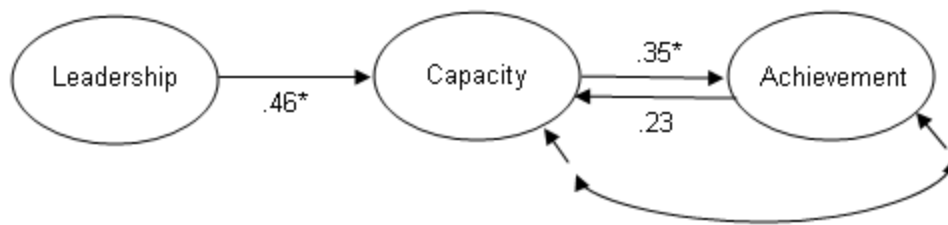


Figure 2. Estimated Unidirectional Effect With Feedback Loop (*p < .05)

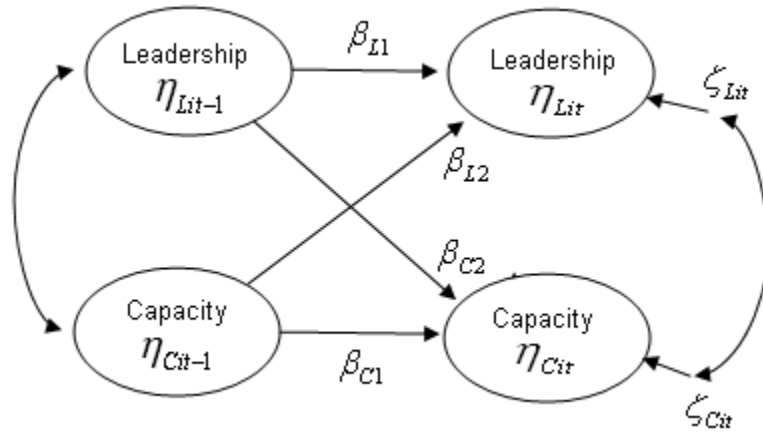


Figure 3. Proposed Mutually-Reinforcing Relationship Over Two Waves of Data

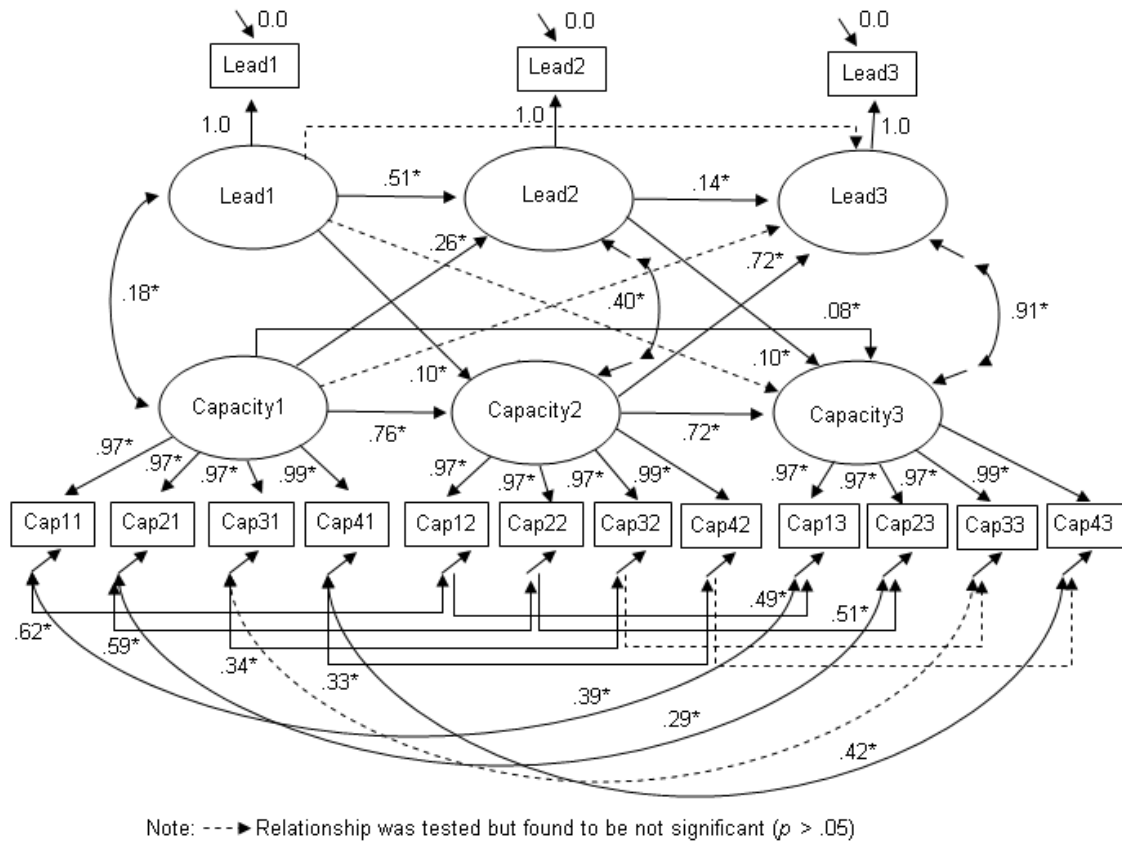
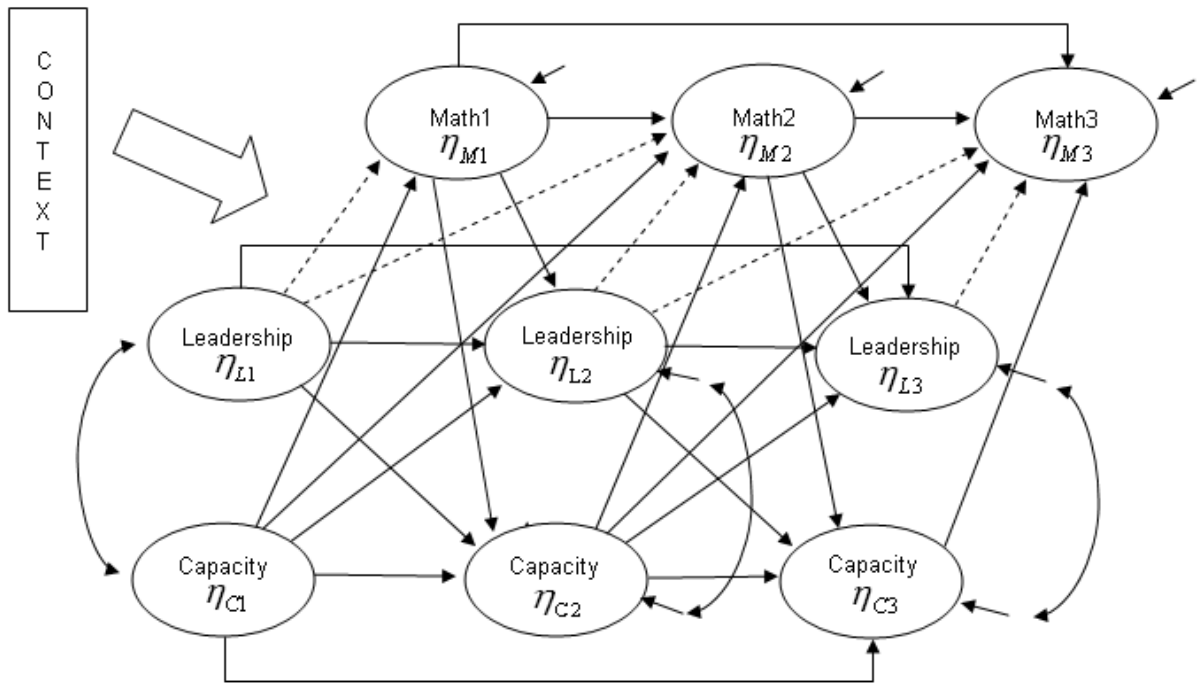


Figure 4. Tested Autoregressive Cross-Lagged Relationship Between Leadership and Improvement Capacity (i.e., with invariant factors and factor loadings) * $p < .05$.



Note: ----> Relationship was tested but proposed to be non-significant.

Figure 5. Proposed Full Cross-Lagged Autoregressive Model.

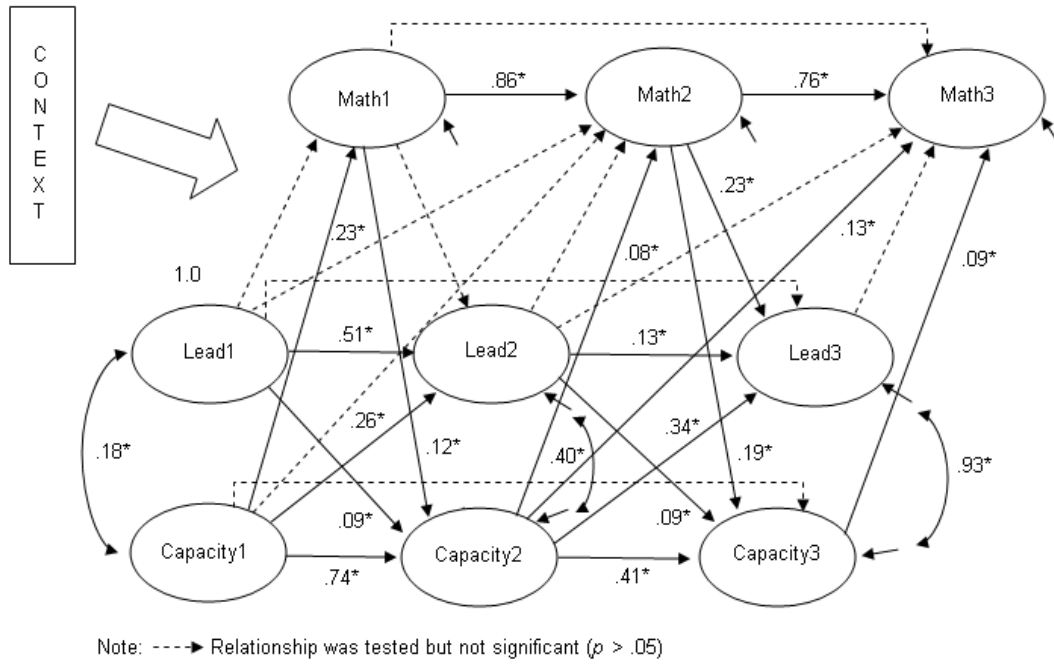


Figure 6. Empirically Derived Mutually-Reinforcing Relationships ($*p < .05$).

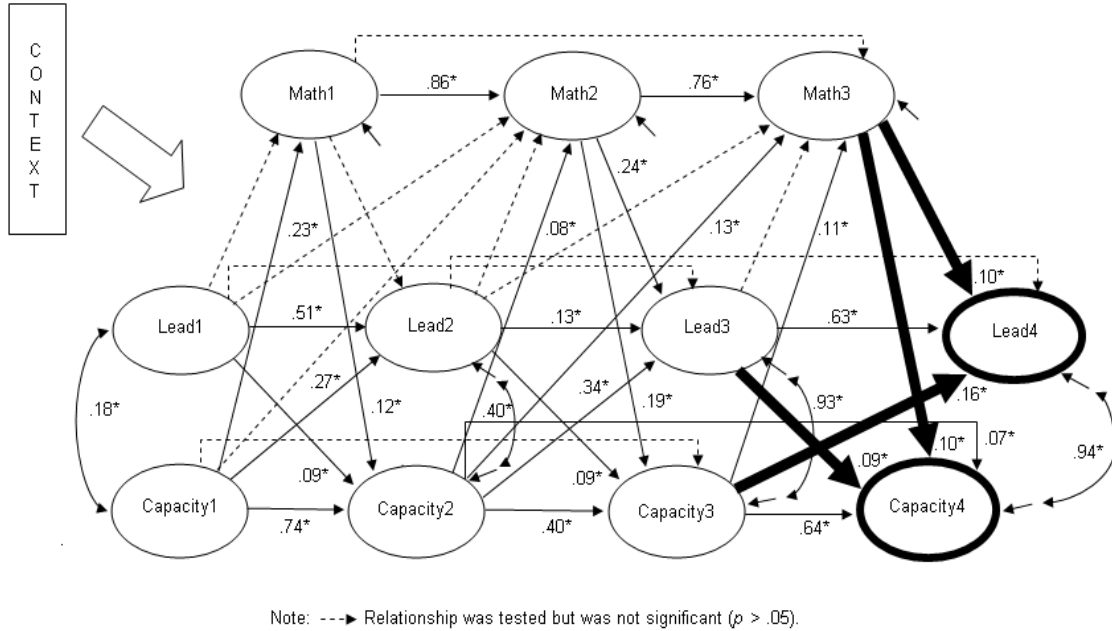


Figure 7. Extending Mutually-Supporting Relationships to Time 4 (* $p < .05$).

Table 1. Model Fit Indices

Model	CFI	SRMR
<i>Preliminary Models</i>		
Measurement Invariance (invariant factors and factor loadings)	0.95	0.04
Measurement Invariance (invariant factors, factor loadings, errors)	0.93	0.05
Model 1 (cross-sectional data with mutual influence)	1.00	0.00
<i>Cross-lagged Models</i> (with invariant factor loadings)		
Model 2 (leadership, capacity, 3 waves)	0.96	0.08
Model 3 (leadership, capacity, achievement, context covariates, 3 waves)	0.95	0.08
Model 4 (constructs, covariates, 4 waves)	0.95	0.08

CFI = Comparative Fit Index; SRMR = Standardized Root Mean Square Residual.

Bios

Dr. Philip Hallinger is the Joseph Lau Chair Professor and Director of the *Asia Pacific Centre for Leadership and Change* at the Hong Kong Institute of Education. His research focuses on school leadership effects, leadership development and problem-based learning. He received his Ed.D. in Administration and Policy analysis from Stanford University. He can be reached at hallinger@gmail.com. His research publications can be accessed at philiphallinger.com

Dr. Ronald Heck holds the Dai Ho Chun Endowed Chair in Education at the U. of Hawaii-Manoa. He received his Ph.D. in Education from the University of California at Santa Barbara. His research focuses on the relationship between school leadership and school improvement. He has also written on research methodology. He can be reached at rheck@hawaii.edu

ⁱ The author would like to acknowledge the useful feedback offered on this manuscript by Edwin Bridges and anonymous reviewers. The authors wish to acknowledge the funding support of the Research Grant Council (RGC) of Hong Kong for its support through the General Research Fund (GRF 840509).

ⁱⁱ Given space constraints and the purpose of this paper, we do not describe the data set from which the example analyses were obtained in great detail. The survey subscales (consisting of 8-10 items each) defining the collaborative leadership and school improvement capacity constructs (with alpha coefficients above 0.80) were developed through confirmatory factor analysis and have been shown in previous studies over a 10-year period to explain levels of school achievement and school growth. Achievement data were obtained from state achievement tests at the individual student level as a series of repeated measures. The tests have been vertically equated to permit examining growth over the three-year period for this student cohort. For a detailed description of the data set, instruments and related psychometric procedures, we refer the reader to Heck & Hallinger, 2009.

ⁱⁱⁱ We used the Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR) to evaluate the models. Values of the CFI near 0.95 and SRMR of 0.08 are often considered evidence of adequate model fit (Raykov & Marcoulides, 2006).

^{iv} The basic measurement model can be defined as

$$y_i = \Lambda \eta_i + \varepsilon_i$$

where Λ a matrix of factor loadings measuring each latent variable η_i , and ε_i are errors associated with items defining each construct which are contained in Θ . Factor variances and covariances are contained in Ψ .

^v 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). At a minimum, the same factor structure and invariant loadings of items on factors should be observed. This analysis is conducted to establish the consistency (i.e., reliability) and validity of our conceptualization of collaborative leadership and school improvement capacity over three measurement occasions. Adequacy of the consistency in measuring these processes simultaneously over time is determined by examining the model fit indices. Once measurement invariance is established, it is possible to examine whether perceptions changed over time. The successive factor means can be simultaneously tested (i.e., with t-tests) against the initial factor mean ($\bar{X}_1 = 0.00$, $SD = 1$), 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., $\bar{X}_2 = 0.05$; $\bar{X}_3 = 0.21$). Although the factor score metric does not reveal the magnitude of the change, the difference between T_1 and T_3 was statistically significant ($t = 3.04$, $p < .01$). We also examined changes in the collaborative leadership factor (which is comprised of one observed scale consisting of eight items). The estimated factor means suggested leadership perceptions were the same between T_1 and T_2 but were not the same between T_1 and T_3 ($t = -2.34$, $p < .05$).

^{vi} Variance in improvement capacity accounted for at Times2-4 was .68, .58, and .71, respectively. Variance in leadership accounted for at Times-2-4 was .38, .47, and .66, respectively. Variance in math accounted for at Times 1-3 was .05, .78, and .46.