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

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

Abstract

Over the past several 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 influenced by features of the organizational environment in which it is enacted. This observation leads us to conclude that the predominant approaches to studying school leadership effects provide an incomplete picture of the relevant processes and paths of influence. In order to address this problem, we examine the potential offered by conceptualizing leadership as a reciprocal or mutual influence process. The paper examines the implications of adopting a reciprocal effects perspective on leadership. It then explores methodological options and issues for investigators to consider in undertaking this rich but challenging approach to understanding how school leadership impacts learning.
Successful 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 reciprocal process 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; Conger & Pearce, 2003; 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 (author id ref.). Most scholars have framed leadership (i.e., either explicitly or implicitly) as an independent variable that drives school change and effectiveness (author id ref.; Bossert, Dwyer, Rowan & Lee, 1982; Bridges, 1970, 1982; Leithwood, Day, Sammons, Harris, & Hopkins, 2006). This approach is, for example, evident in the most frequently cited mediated-effects studies of school leadership effects in the published literature (e.g., author id. ref.; author id. ref.; Goldring & Pasternak, 1994; Krüger, Witziers, & Sleegers, 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 and student academic achievement (e.g., author id. ref.; Bossert et al., 1982; Leithwood et al., 2006; Pitner, 1988). The causal ordering issue refers specifically to the question whether conclusions about ‘effective school leadership’ reflect valid measurement of leadership impact on the organization, or optimistic attributions based on associations between leadership and organizational
performance (author id. ref; Bridges, 1982; Rowan et al., 1982). We note that virtually all quantitative studies of leadership and learning conclude with such caveats.

One reason for this derives from the predominance of cross-sectional research in this field. Unfortunately, cross-sectional research designs are poorly suited to resolving the complex issue of causal ordering of variables in models that seek to link leadership to school effectiveness (Witziers et al., 2003). This limitation becomes even more evident if the research seeks to investigate the impact of leadership on school improvement, a process which by definition unfolds over time (author id.; Day, Sammons, Leithwood, Hopkins, Harris, Gu, & Brown, 2010; Mulford & Silins, 2009). Thus, we contend that the causal ordering of leadership and other relevant variables that might contribute to student learning represents an important theoretical issue with practical implications.

Mutual influence or reciprocal effects models offer an alternative means of unpacking the issue of causal ordering. Rather than conceptualizing leaders as the either the ‘origin or the pawn‘ (Bridges, 1970) administrative decision-making, a reciprocal-effects model would propose bi-directional rather than unidirectional effects (author id. ref.; Pitner, 1988; Tate, 2008). While reciprocal-effects models have been discussed by scholars for several decades, there have been 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., author id. ref.; Mulford & Silins, 2003, 2009).

In this paper, we explore the potential of reciprocal-effects modeling in studies of school leadership effects on learning. We begin by discussing the conceptual basis...
for this approach. Then we discuss relevant methodological requirements and options. Testing reciprocal-effects models through the use of longitudinal data raises a host of new challenges, with relatively few published studies to serve as models (e.g., Marsh & Cravens, 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 requirements and options, but also to explicate these through the analysis of an illustrative longitudinal data set. We hope that this approach will provide a useful foundation for other scholars who see benefit in the exploring the potential of mutual influence models in the study of leadership for learning.

Conceptualizing School Leadership Effects

The term leadership for learning (e.g., MacBeath & Cheng, 2008) implies that causal linkages exist between the actions of leaders and school learning outcomes. The intellectual lineage for studies of leadership and learning traces back to initial explorations by scholars in the USA during the 1960s (e.g., Bridges, 1967; Gross & Herriot, 1965). As time passed, scholars sought to assess whether school leadership effects on learning could be detected through quantitative research (author id. ref.; Bossert et al., 1982). Subsequently, it has sought to identify the paths through which leadership impacts school performance (author id. ref.; Leithwood, Anderson, Mascall, & Strauss, in press; Leithwood et al., 2006, in press; Robinson, Lloyd, & Rowe, 2008).

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). Over time, scholars became increasingly critical of the capacity for these over-simplified models to
capture the complexities inherent in leadership for learning (author id. ref.).

Consequently, researchers began to explore more complex models that conceptualized a variety of moderating and mediating variables that impact leadership and learning (e.g., author id. ref.; Leithwood & Jantzi, 1999, 2000; Wiley, 2001).

In recent years influential scholars internationally (author id. ref.; (Leithwood et al., in press; Mulford & Silins, 2009) have argued forcefully that the ‘next generation’ of leadership for learning studies should aim at exploring the avenues and paths by which leadership effects on learning accrue. Mediated and reciprocal effects models are foremost among those that seek to explore these paths (see Models B and C in Figure 1). Thus, this section of the paper is organized around an elaboration of these perspectives on leadership and learning.

**Mediated Effects of Leadership on Learning**

In recent years, a large body of international research supports the view that school leadership can have significant *indirect* effects on student learning (author id. ref.; Bell, Bolam, & Cubillo, 2003; Bossert et al., 1982; Leithwood et al., 2006, in press; Robinson et al., 2008; Wiley, 2001; Witziers, Bosker, & Kruger, 2003). This perspective is represented by Model B in Figure 1. Scholars increasingly embrace 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 (author id. ref.; Leithwood et al., 2006, in press; Robinson et al., 2008). Empirical evidence, though not conclusive, does provide insight into these *paths* by which leaders impact teaching and learning. Specifically, research indicates that school improvement leadership:

- Impacts conditions that create positive learning environments for students (author id. ref.; Leithwood et al., 2006, in press; Robinson et al., 2008;
This description of the means by which leadership impacts school improvement is consistent with what scholars have termed a mediated-effects model of leadership (Baron & Kenny, 1986; Pitner, 1988). Leadership effects on learning are produced indirectly through their impact on people, structures, and processes in the school that are more proximal to students (author id, ref; Bossert et al., 1982; Leithwood et al., in press). While this conceptualization of leadership represents an advance over earlier two-factor studies, mediated-effects models as portrayed by Model B in Figure 1 still frame leadership as the cause of change in the organization. Such models have been critiqued as asserting a heroic role for leaders that fail to recognize the systemic forces and constraints under which leaders operate (Bossert et al., 1982; Bridges, 1970; 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

- Mediates academic expectations embedded in curriculum standards, structures, and processes as well as the academic support that students receive (author id. ref.; Hill & Rowe, 1996; Robinson et al., 2008).
- Employs improvement strategies that are matched to the changing state of the school over time (author id. ref.; 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 (Leithwood et al., 2006, in press; Robinson et al., 2008).
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, we note that the mediated-effects studies have generally relied on cross-sectional data describing leadership and related school-level variables. Scholars have noted important limitations that arise from an over-reliance on cross-sectional data to assess the effects of leadership (author id. ref.; Ogawa & Bossert, 1995). For example, Hallinger and Heck (1996) earlier observed:

[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. Interpretation of data from correlational studies of principal effects is still hindered by the absence of longitudinal research, both quantitative and qualitative. (p. 36)

Most specifically, the limitations of cross-sectional research designs include problems with capturing the complete set of interactions among variables and determining the direction of causality of proposed relationships. These problems further compound when the dependent variable of interest is school change or improvement, both of which imply the need to measure leadership impact over time. Both theorists and practitioners know that school life is complex and cyclical (e.g., days within each week, weeks within each month, months within each year, years within a period of stability or change). Understanding how the effects of school processes unfold therefore calls for data and research designs that can incorporate some of this complexity within the analysis. Recognition of the limitations of cross-sectional snapshots of school processes in school leadership research led scholars to propose alternative ways of understanding how leadership might impact learning.
This is the pre-published version.

(author id., ref; Pitner, 1988; Southworth, 2002). One alternative model formulation emphasizes reciprocal relationships. We summarize this type of mutual causation as Model C in Figure 1 with double arrows between major constructs (A ▷ B).

Reciprocal Effects Models

School improvement is, by definition, a process that involves change in the state of the organization over time. This observation suggests that the empirical study of school improvement leadership requires models that take into account changing relationships among relevant variables over time. As Hallinger and Heck (1996) 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” (p. 19).

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; Williams & Podsakoff, 189). 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 et al., 1983; Rao & Mawhinney, 1991; Schriesheim & Hinkin, 1990), and distributed leadership (Gronn, 2002; Spillane, 2006). Only in the last decade, however, has recognition of the interactive 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). We note that, to date, there has only been a limited effort to model reciprocal effects of leadership and learning in education quantitatively (author id. ref.; Mulford & Silins, 2003).
We suggest that progress in testing conceptual models that incorporate reciprocal causation in leadership research has been hindered by several methodological challenges. Griffin (1997) noted some of these obstacles in his review of the literature on leadership effects.

Reciprocal interaction has been incorporated into a variety of organizational behavior theories... However, there is a frequent tone of pessimism that such an interaction can be captured by anything except the most complex research enterprises. Kozlowski and Doherty's (1989) integration of climate and leadership concepts included the comment that it was “too early to specify and test reciprocal effects” (p.552). George and Brief's (1992) review of individual organizational spontaneity and group affect disclaimed “available data are too sparse...but, on conceptual, grounds it probably could be argued that the relationship is reciprocal” (p.321). More broadly, the estimation of reciprocal effects has been methodologically problematic in organizational research (Williams & Podsakoff, 1989). (Griffin, 1997)

In models of reciprocal interaction there is an explicit assumption that behavioral adaptation unfolds over time (Griffin, 1997; Marsh & Cravens, 2006; Tate, 2008). Longitudinal data that describe change in organizational processes over time are, however, difficult to obtain, especially on a scale sufficient to assess the effects of leadership across comparable organizational units. Moreover, researchers have not always utilized analytical tools capable of modeling reciprocal effects adequately over time (author id. ref.; Griffin, 1997; Marsh & Cravens, 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 (author id. ref.; Hill & Rowe, 1996). Despite these challenges, we cannot overstate the importance of using
longitudinal designs in leadership research where progress has both theoretical implications and practical utility (Griffin, 1997; Ogawa & Bossert, 1995; Tate, 2008).

As suggested by Bridges (1970, 1977), there is a strong theoretical bias towards viewing leaders as actors or agents rather than pawns. Second, there has been a methodological bias toward cross-sectional data, since it is a major challenge to obtain and analyze longitudinal data that are able to capture the effects of mutual influence processes within multilevel school settings. In short, most state or national data bases are not routinely organized in this manner, though this is changing with increased policymaker interest in monitoring student growth in learning. We believe that the conceptual arguments for studying reciprocal relationships are strong.

We have previously emphasized that holding school leaders accountable for student outcomes, while politically attractive, has been based on scant empirical data that directly links leadership efforts to improved student outcomes—that is, most evidence of this supposed direct effect is through relationships established at one point in time where it is impossible to determine whether better leaders do indeed influence achievement outcomes positively or merely select school settings with stronger achievement (author id. ref.). In this study, for example, we noted only about one-third of school principals stayed at their school over the four-year period of our study. Yet having a stable school principal was not directly related to growth in student learning in either reading or math.

This suggests a much more complex process underlying the dynamics of school leadership and school improvement processes as they unfold over time—processes that certainly extend beyond the heroic efforts of single individuals acting as causal agents within their schools. For this reason, we look at leadership for learning as comprised of processes that build organizational capacity—that is, the
capacity for shared, or collaborative, leadership as well as the capacity for improving the school’s academic structures and culture (e.g., Barth, 2001; Fullan, 2006; Ogawa & Bossert, 1995). Following Pitner’s (1988) conceptualization, we surmised that a ‘cross-lagged’ model of reciprocal influence, which measures organizational and outcome processes over a period of time, would be a potentially more powerful means of capturing the effects of leadership in organizations than analyses conducted at single points in time. This proposition provides a transition into the methodological challenges of employing reciprocal effects models in studies of leadership for learning.

Methodological Issues in Examining Reciprocal Effects

In the previous section of this paper, we outlined the conceptual rationale for employing reciprocal-effects models in leadership research. In this section, we lay out the basics of conducting a cross-lagged longitudinal analysis of collaborative leadership effects on school improvement processes. However, it is, in fact, difficult to elaborate on the conceptual issues without reference to how one might operationalize these models. With this in mind, we seek both to illustrate and explore the options for modeling reciprocal relationships between leadership and relevant organizational variables in a school environment. Our approach in this section is to develop and test one particular type of reciprocal-effects model. It will quickly become apparent that the conceptual and methodological issues involved in studying reciprocal effects are closely intertwined.

Because organizational settings are not readily amenable to experimental manipulations, research examining temporal relationships is often conducted with longitudinal panel studies in which variables are measured on at least two occasions. The panel approach dates at least from Lazarsfeld and Fiske (1938), who 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 was that it did not rely on instantaneous relationships between simultaneously measured variables (Oud, 2002).

**Methodology Employed in Testing Illustrative Models**

In this section of the paper, we illustrate methodological issues that arise in the use of reciprocal effects models through reference to a longitudinal data set that describes leadership and learning in schools. The data set consists of survey data collected from teachers in 198 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).

Within any given year, survey data from schools on collaborative leadership processes and school capacity for improvement were collected *before* student achievement data. The temporal sequence of data collection, therefore, makes the available data ideal for testing this type of proposed cross-lagged 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.

We note that structural equation modeling (SEM) is 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. 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-effects 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

There are a number of different ways to model mutual influence utilizing either cross-sectional or longitudinal data (Kline, 2004). One way is through cross-lagged longitudinal modeling. Such 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 & Cravens, 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”) relationship over two periods of time is defined as A1 → B2 and B1 → A2. School leaders may, for example, 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.

Cross-lagged longitudinal modeling is most advantageously used to explore causal relationships between two (or more) variables over time. It is, for example, possible that (1) A → B (i.e., leadership behavior causes change in subsequent school improvement capacity); (2) B → A (i.e., school improvement capacity causes subsequent change in leadership behavior); or whether a more complex set of mutually-reinforcing causal relationships may exist between the variables. This third possibility is which notated as (3) A → B (i.e., prior leadership causes subsequent changes in improvement capacity and prior improvement capacity causes subsequent leadership changes). Thus, a reciprocal-effects 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 reciprocal-effects model using cross-sectional data. Central to the understanding of reciprocity or mutual influence is the notion of feedback between variables over time. A feedback
loop 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 & Cravens, 2006).

If one proposes the existence of a feedback loop between two variables, this implies more than a simple cause and effect (A→B) relationship from the independent variable (A) to the dependent variable (B). It also implies the converse—that B affects A (B → A). The presence of a feedback loop (A ↔ B) 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 a reciprocal relationship is not the same as a two-headed arrow (A ↔ B). This would simply indicate a covariance (or correlation) between the two variables. It does not necessarily imply any temporal ordering of the effects and therefore fails to capture systemic feedback between variables as their interaction unfolds over time.

In Figure 2, we provide the results of our proposed reciprocal-effects model with a single wave of data from our study. The model fits the data perfectly (see Table 1), since in this simple example there are just as many paths estimated as available information in the school-level covariance matrix. For ease of presentation, we only focus on relationships between the latent constructs and not the measurement model which defines their observed indicators or measurement errors.iii 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 result 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. We could, of course, extend the logic of reciprocal relationships by suggesting that the state of the school’s improvement capacity might also influence leadership actions, which would imply addition of another path to the proposed model (i.e., provided we could meet the requirements for model identification). As shown in the figure, the errors in equations involved in feedback loops (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). This is consistent with the logic of reciprocity, since if we assume that A and B mutually influence each other, then we may reasonably expect that they have common omitted causes (Anderson & Williams, 1992; Kline, 2004).

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. That is, one must assume that the relationship specified in the 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 is often difficult to justify. This leads to the need to explore reciprocal relationships 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, the limitations of this approach are considerable. As noted above, the primary limitation concerns the requirement that the system is stable. 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; Tate, 2008).

Ogawa and Bossert (1995) succinctly 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. A more robust way to specify reciprocal effects lies in the use of longitudinal panel designs where variables are each measured on two or more different occasions (Bollen & Curren, 2006; Lazarsfeld & Fiske, 1938; Marsh & Cravens, 2006; Williams & Podsakoff, 1989). More specifically, we suggest that autoregressive cross-lagged longitudinal models offer potential 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 (Klein, 2004). Such models cannot be analysed by ordinary least squares (OLS) regression due to correlated errors between observations.

*Specifying a cross-lagged longitudinal model.* As we have emphasized, one key aspect of longitudinal reciprocal-effects models is the testing of proposed mutually-reinforcing relationships. We first develop an autoregressive cross-lagged model to investigate the possibility that collaborative 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, once again in the figure we do not show the measurement model, which consists of the observed indicators of each latent variable, nor the correlated residuals for each observed indicator across time occasions (see Figure 4 for the complete results).
The model is based on the assumption that each latent construct \( \eta \) 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 \( \eta_{L,t-1} \) in the figure to indicate the measurement of the latent leadership construct at Time 1 or T1) affects the subsequent state of B (which we will label \( \eta_{C,t} \) to indicate the measurement of the latent improvement capacity construct at Time 2 or T2) and, simultaneously, an earlier state of improvement capacity (\( \eta_{C,t-1} \) at T1) affects the later state of leadership (\( \eta_{L,t} \) at T2). 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 (Schluetler 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.

An initial (T1) covariance (or correlation) between the constructs is also proposed in Figure 3. There is also a covariance proposed (curved two-headed arrow) between the residuals 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 (which are referred to as autocorrelated measurement errors). Failure to account for covariation between errors can bias the
model estimates of reciprocal effects (Sturgis et al., 2004; Williams and Podsakoff, 1989).

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:

\[
\eta_{Li} = \alpha_{Li} + \beta_{Li} \eta_{Li-1} + \beta_{Ci} \eta_{Ci-1} + \zeta_{Li},
\]

\[
\eta_{Ci} = \alpha_{Ci} + \beta_{Ci} \eta_{Ci-1} + \beta_{Li} \eta_{Li-1} + \zeta_{Ci},
\]

(1)

where \( \alpha_{Li} \) and \( \alpha_{Ci} \) are intercepts, \( \eta_{Li-1} \) and \( \eta_{Ci-1} \) represent the latent leadership and improvement capacity constructs for individual \( i \) at time \( t-1 \), \( \beta_{Li} \) and \( \beta_{Ci} \) are autoregressive structural coefficients, \( \beta_{Li} \) and \( \beta_{Ci} \) are cross-lagged structural coefficients, and \( \zeta_{Li} \) and \( \zeta_{Ci} \) are errors in predicting each outcome at Time 2.

The magnitude of the cross-lagged coefficients indicates how much variation in \( \eta_{Li-1} \) predicts aggregate changes in \( \eta_{Li} \), 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 T3 (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., Rogosa, 1980; Oud, 2002). A key limitation is that in discrete-time models different time lags can result in different estimates of effects. This problem can, however, be alleviated by continuous-time modeling (e.g., Oud, 2002).

The cross-lagged approach is also not as adaptable to flexible treatments of time as, for example, latent growth models, which can provide individually-varying trajectories across organizational settings and can incorporate nonlinear growth.
(Sturgis et al., 2004). If time intervals are poorly chosen in cross-lagged models, it is possible to miss the effect because the interval chose was either too short or too long (so the effect has faded). Since no one lag is sufficient to understand a 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 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, which provides an average effect across the period of time under consideration (Sturgis et al., 2004).

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).iv We rejected the hypothesis of equal errors for observed items across time, however (not tabled). We also note in passing that we could also represent cross-lagged effects with at least three measurement occasions as a latent curve (i.e., growth) model (author id. ref.). 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 find that it fits the data quite well (see Table 1). Figure 4
summarizes the invariant factor loadings across the three measurement occasions for each of the four indicators of school improvement capacity. As shown in the figure, the errors for each indicator defining improvement capacity were not found to be 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).

The cross-lagged relationship proposed between leadership at T1 and improvement capacity at T2 is significant ($\beta = .09, p < .05$), as is the relationship between improvement capacity at T1 and leadership at T2 ($\beta = .26, p < .05$). The cross-lagged relationships are also similar between T2 and T3 (leadership $\rightarrow$ capacity $\beta = 0.10, p < .05$; capacity $\rightarrow$ leadership $\beta = 0.72, p < .05$). The stability relationships between leadership at T1 and T2 (standardized $\beta = 0.51$) 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 ($\beta = .13, p < .05$). This weaker stability coefficient corresponds with the fact that leadership perceptions on average dipped slightly between T1 and T2. 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 0.66 (at T3, the correlation is stronger at 0.91). These results provide support for the view that leadership and improvement capacity are mutually reinforcing over time.
We can also note that previous improvement capacity appears moderately related to subsequent leadership, while previous leadership is only weakly related to subsequent improvement capacity.

**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 to the model (see Figure 5). The model conceptualizes leadership as both an independent and dependent variable embedded within the organizational context; that is, leadership is still viewed as impacting school improvement in learning primarily through *indirect* paths (author id. ref.). 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 cross-lagged effect of previous leadership on subsequent achievement (at T1, T2, and T3) with *dotted* lines in Figure 5 to emphasize that we expect the paths to be non-significant, but we test for the presence of a *direct* effect at each occasion. For ease of presentation, once again we also do not show the observed indicators and error terms for each measurement occasion.

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 (assuming equilibrium in the system).

Insert Figure 5 about Here

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 (author id. ref.; Ogawa & Bossert, 1995). We can test these contingencies by holding the cross-lagged and stability paths to be invariant over the measurement occasions and examining the change in chi-square coefficient ($\Delta \chi^2$) associated with the degrees of freedom in each specific test of invariance. If the change in chi-square is small when the parameters of interest are held invariant, it suggests we can accept the relationships are the same over time. This implies an “average” effect for the pattern of relationships over the time of the study (Sturgis et al., 2004). 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.

*Testing the full model.* When we test the model with unrestricted autocorrelation (stability) and cross-lagged effects, we find that it fits the data well using typical SEM fit indices (see Table 1). The results 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.

We first note that the mutually-reinforcing relationship between leadership and capacity building is again supported over the three occasions of measurement. Second, although tested, we note that prior leadership was not directly related to subsequent achievement at each measurement occasion. In contrast, prior
improvement capacity was related to subsequent achievement at each occasion (with coefficients ranging from 0.08 to 0.23, \( p < .05 \)). We also note 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 (\( p > .05 \)), but it was predictive of leadership at T3 (\( \beta = 0.23, p < .05 \)).

Testing the invariance of the autocorrelation and cross-lagged effects. We next tested whether the stability effects (i.e., autoregressive paths for each construct) were the same for each occasion. We found that we could not consider these effects as invariant (\( \Delta \chi^2 = 40.4, 6 \text{ df}, p > .10 \), not tabled). We then tested whether the cross-lagged effects were invariant over time. We found that we could not consider the cross-lagged effects to be invariant (\( \Delta \chi^2 = 50.2, 8 \text{ df}, p > .10 \), not tabled).

Finally, only a limited number of covariates in the model (i.e., student composition, staff stability, teacher professional qualifications) exerted small effects (\( \beta < .25, p < .05 \)) on the constructs on one or more measurement occasions. Including them, however, improves the examination of the model’s component constructs.

Insert Figure 6 about Here

Extending the Reciprocal-Effects Model

Empirical support in Figure 6 for the mutual-reinforcing model supports the theoretical premise that changes in organizational leadership and school improvement capacity represent a mutual growth process (Muthén & Muthén, 1998-2006). 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 our proposed model in Figures 2-6 by examining whether it is useful in explaining the variation in the leadership and capacity constructs at a fourth point in time (i.e., using school survey data that became available after our primary model was developed and tested).

When we examine this final model (see Figure 7), we find that leadership at T3 was related to capacity at T4 ($\beta = 0.09, p < .05$) and vice versa ($\beta = 0.16, p < .05$). The stability coefficients for each construct were also strong between T3 and T4 ($\beta > 0.60, p < .05$). This suggests that leadership was more stable between T3 and T4 than between T2 and T3. Overall, the model coefficients at T4 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. Moreover, consistent with our second premise, we found that achievement at T3 was positively related to subsequent changes at T4 in both leadership and capacity building.

We also tested the invariance of the stability coefficients over the four time periods and found them invariant or not invariant for all three constructs ($\Delta \chi^2 = 40.4, 6 \text{ df}, p > .10$, not tabled). Finally, we also tested whether the cross-lagged effects were invariant over the four time periods. We found that we could not consider the cross-lagged effects to be invariant ($\Delta \chi^2 = 50.2, 8 \text{ df}, p > .10$, not tabled).

Our results therefore show a positive longitudinal effect of achievement levels in subsequent levels of both capacity building and provide some support for achievement on collaborative leadership, especially during the latter stages of the temporal sequence examined (i.e., Time 3 and Time 4). These results provide support
for the premise that schools can improve their outcomes regardless of their initial achievement levels.

Conclusions

This paper is located within the intellectual lineage of research that studies school leadership effects on learning (Bossert et al., 1982; 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 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, Patten, & Jantzi, 2010, p.1)

The body of empirical research in school leadership effects on learning has progressed from relatively simple towards more complex models. 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). Although subsequent reviews suggest that researchers have generally heeded this recommendation, most subsequent studies have employed mediated-effects models (author id. ref.; Leithwood et al., 2010; Robinson et al., 2008). We were only able to locate two studies of school leadership effects that employed reciprocal-effects models (author id. ref.; Mulford & Silins, 2009).

This paper has sought to examine the conceptual basis for employing reciprocal-effects models of leadership and learning and to demonstrate alternative approaches to examining them. The concept of mutual influence has long been
acknowledged in the leadership literature, yet seldom studied empirically (Griffin, 1997). The illustrations of reciprocal-effects modeling provided in the paper support the conclusion that this is a viable alternative means of understanding how leadership contributes to learning.

We note our results should be considered along with several limitations. Because of their increased complexity, the cross-lagged models tested can be generally more difficult to estimate. They often require very accurate starting values to help minimize the likelihood function. Moreover, it becomes essential to examine proposed solutions for illogical parameters (e.g., negative error variances, large standard errors). Choice of discrete time intervals has been shown to influence the pattern of results. It is also important to acknowledge that such models may still not resolve issues of whether variable A causes B or variable B causes A unless the criteria previously outlined can be met. As we noted, one condition is the stability of the causal structure; that is, the structure of proposed relationships does not change over time.

Examining our model over four time periods, however, provides support for the assumption that the structure of the proposed relationships was stable over time. Note that this could only be assumed, but not be verified in a cross-sectional test of the model. It is still important to note, however, that even this type of test does not provide complete protection against a possible selection-bias argument. For example, teachers may perceive improvement capacity more positively in schools that achieve at high levels over long periods of time.

In addition, caution must be exercised in using SEM applications to test substantive theories. Omitted variables are common sources of misspecification that can produce misleading results (Klein, 2004). We did include a full range of school
context variables in our model investigations, but it is likely that there are other educational processes that influence student learning. Some of these include teacher classroom behavior, student grouping strategies, and student academic and social integration within the school.

Despite these limitations, our results provide initial empirical support for contingency views of leadership. Indeed, we view the application of reciprocal-effects models as highly complementary to the body of unidirectional, mediated-effects studies. We encourage other researchers working in this important area of research to employ both types of models as a means of clarifying and expanding our understanding of the relationship between leadership and learning.
References


Table 1: Model Fit Indices

<table>
<thead>
<tr>
<th>Model</th>
<th>CFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement Invariance (i.e., invariant factors and factor loadings over time)</td>
<td>0.97</td>
<td>0.07</td>
</tr>
<tr>
<td>Models Tested</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1 (Cross-sectional)</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Model 2 (Leadership, capacity)</td>
<td>0.96</td>
<td>0.09</td>
</tr>
<tr>
<td>Model 3 (3 constructs)</td>
<td>0.95</td>
<td>0.08</td>
</tr>
<tr>
<td>Model 4 (3 constructs, 4 occasions)</td>
<td>0.95</td>
<td>0.08</td>
</tr>
</tbody>
</table>

CFI=Comparative Fit Index; SRMR = Standardized Root Mean Square Residual.
Figure 1: Modeling School Leadership Effects on Learning

*adapted from Pitner, 1998, pp. 105-108*
Figure 2:
Estimated Unidirectional Effect and Feedback Loop (*p < .05)
Figure 3: Proposed Mutually-Reinforcing Relationship Over Two Waves.
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 found to be not significant (p > .05)
Figure 5: Proposed Full Cross-Lagged Model.
Figure 6: Empirical Mutually-Reinforcing Relationships (*p < .05)
Figure 7:
Extending the Mutually-Supporting Relationships to Time 4 (* p < .05).

Note: --> Relationship was tested but was not significant (p > .05)
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).

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 Author, 2009.

The basic measurement model can be defined as

\[ y_i = \mu + \lambda \eta_i + \epsilon_i, \]

where \( \mu \) is a vector of intercepts, \( \lambda \) a matrix of factor loadings measuring each latent construct \( \eta_i \), and \( \epsilon_i \) are errors associated with items defining each latent construct.

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., \( \bar{X}_2 = 0.07; \bar{X}_3 = 0.09 \)). Although the factor score metric does not reveal the magnitude of the change, the difference was statistically significant (\( t = 4.83, p < .01 \)). 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 not the same over time (\( t = 2.34, p < .05 \)).