Modeling the longitudinal effects of school leadership on teaching and learning

Ronald H. Heck
Department of Educational Administration, University of Hawai'i, Manoa, Honolulu, Hawai'I, USA, and
Philip Hallinger
Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education, Hong Kong, China

Received 31 August 2013
Revised 3 March 2014
Accepted 26 March 2014

Abstract

Purpose – The purpose of this paper is to test a multilevel, cross-classified model that seeks to illuminate the dynamic nature of relationships among leadership, teaching quality, and student learning in school improvement. The study's primary goal is to shed light on the paths through which leadership influences student learning. At the school level, the model examines the mediating effect of the school's instructional environment on leadership and student learning. At the classroom level, it examines how instructionally focussed leadership can moderate teacher effects on student learning. Then these multiple paths are examined in a single model that seeks to test and highlight the means by which leadership contributes to school improvement.

Design/methodology/approach – The current study employed a multilevel longitudinal data set drawn from 60 primary schools in one state in the USA. Using a cross-classification approach to quantitative modeling, the research analyzes the complex cross-level interactions that characterize school-level and classroom level practices that contribute to school improvement and student learning.

Findings – The results illustrate the utility of specifying multilevel relationships when examining the "paths" that link school leadership to student learning. First, leadership effects on student learning were fully mediated by the quality of the school's instructional environment. Second, the findings indicated that the classroom-related paths examined in this study directly influenced the measures of student math achievement. Third, the research found that instructionally focussed school leadership moderated the effect of individual teachers on student learning. Fourth, the results suggest that school leaders can enhance student outcomes by creating conditions that lead to greater consistency in levels of effectiveness across teachers.

Practical implications – The study makes substantive contributions to the global knowledge base on school improvement by testing and elaborating on the "paths" that link school leadership and student learning. More specifically, the findings offer insights into strategic targets that instructional leaders can employ to enhance teacher effectiveness and school improvement. Thus, these results both support and extend findings from prior cross-sectional and longitudinal studies of leadership and school improvement.

Originality/value – This is the first study that has tested a conceptualization of leadership for learning in a single "cross-classified longitudinal model" capable of capturing interactions among leadership, classroom teaching processes and growth in student learning. The research illustrates one "state-of-the-art" methodological approach for analyzing longitudinal data collected at both the school and classroom levels when studying school improvement.

Keywords Leadership, Effectiveness, School improvement, Student learning, Teaching quality, School leadership, School change, Effects of leadership, Instructionally focussed leadership

Paper type Research paper

Do teachers make a difference? Can leaders enhance the quality of teaching and learning in schools? These questions have consumed the attention of policymakers, school practitioners and researchers over the past half-century as governments have sought more systematic strategies for improving the quality of schooling

Journal of Educational Administration
Vol. 52 No. 5, 2014
pp. 653-681

0957-8234
DOI 10.1108/JEA-08-2013-0097

(Creemers and Kyriakides, 2008; Erickson, 1967; Hallinger, 2011; Teddlie and Stringfield, 1993; Reynolds *et al.*, 2000). In response, a large body of international research has accumulated which supports the view that both quality of school leadership and teaching can have a significant impact on student learning outcomes (Bossert *et al.*, 1982; Coelli and Green, 2012; Creemers and Kyriakides, 2008; Erickson, 1967; Hallinger, 2011; Hallinger and Heck, 1996, 1998; Leithwood *et al.*, 2004, 2006; Robinson, 2011; Robinson *et al.*, 2008; Urick and Bowers, 2011; Witziers *et al.*, 2003). Scholars increasingly accept the assertion that the impact of leadership on learning is achieved indirectly by shaping conditions that contribute to effective teaching and learning (Creemers and Kyriakides, 2008; Hallinger, 2011; Hallinger and Heck, 1996, 1998; Leithwood *et al.*, 2004, 2006; May and Supovitz, 2011; Robinson *et al.*, 2008; Sebastian and Allensworth, 2012; Southworth, 2002; Urick and Bowers, 2011). Consequently, over the past 15 years, scholarship in this domain has focussed primarily on identifying and testing the "paths" through which leadership influences student learning in schools (Hallinger and Heck, 1996, 1998, 2011; Heck and Hallinger, 2009, 2010; Leithwood *et al.*, 2004, 2006, 2010a, b; Marks and Printy, 2003; Mulford and Silins, 2009; Printy *et al.*, 2009; Robinson *et al.*, 2008; Witziers *et al.*, 2003).

Limitations in the conceptual models and research designs used in this research domain have, however, continued to impede efforts to illuminate fully the paths through which leadership impacts student learning and school improvement (Bossert *et al.*, 1982; Hallinger and Heck, 1996, 2011; Reynolds *et al.*, 2000). Although "change in teacher practice and effectiveness" is often included as a focal variable in conceptualizations of leadership for learning, it has been a challenge for researchers to examine change in leadership, teacher practice, and student learning within a single study. Although most researchers assume that leadership impacts student learning by enhancing the quality of teaching, we are unaware of any studies of leadership effects on student learning outcomes that have included measures of classroom teaching. Instead, researchers have tended to study other school-level constructs (e.g. school climate, academic capacity, academic optimism, organizational learning, collective efficacy, school culture) as mediators of leadership effects on student learning (e.g. Day *et al.*, 2011; Geijsel *et al.*, 1999; Hallinger *et al.*, 1996; Hallinger and Heck, 2011; Heck and Hallinger, 2010; Kruger *et al.*, 2007; Leithwood and Jantzi, 2000; Marks and Printy, 2003; Opdenakker and Van Damme, 2007; Wiley, 2001).

Another seldom tested assumption among these studies is that leadership effects on teaching, learning and school improvement unfold over time (Geijsel *et al.*, 2003; Heck and Hallinger, 2009, 2010; Ko *et al.*, 2012; Mulford and Silins, 2003, 2009; Reynolds *et al.*, 2000). Indeed as several scholars have demonstrated (Day *et al.*, 2011; Hallinger and Heck, 2011; Mulford and Silins, 2009), it can take several years for patterns of school improvement to become visible, especially when examining the aggregated performance of many schools.

Consistent with the purpose of this special issue, we frame the assertion that school improvement processes are mediated by time as an "assumption" that shapes the requirements for high impact research in this domain. Notably, to date, most research on school improvement has consisted of case studies and cross-sectional surveys (Hallinger and Heck, 2011; Ko *et al.*, 2012; Luyten *et al.*, 2005; Mulford and Silins, 2003; Stoll and Fink, 1996). Case studies are useful for describing patterns of change in school processes "up close" and elaborating on the meaning of quantitative results. However, they lack the broader applicability and conclusiveness necessary needed to guide policy discourse. Although researchers sometimes use cross-sectional surveys of school performance data to reconstruct "what happened" prior to the snapshot, conclusions drawn from this kind of research design remain quite speculative (e.g. Day *et al.*, 2011; Walker and Ko, 2011). Thus, we assert that case study and cross-sectional research designs cannot, by themselves, build a sound knowledge base in school improvement research, policy and practice (Creemers and Kyriakides, 2008; Heck and Hallinger, 2009; Luyten *et al.*, 2005; Reynolds *et al.*, 2000).

Recently, concerns over these limitations of the existing knowledge base have led scholars toward the application of quantitative, longitudinal research designs in studies of school improvement (e.g. Hallinger and Heck, 2011; Heck and Hallinger, 2009, 2010; Mulford and Silins, 2009; Thoonen *et al.*, 2012). Longitudinal research designs enable explicit specification and testing of the "temporal ordering" of relevant organizational processes, thereby shedding new light on the nature of causal relationships. This new generation of school improvement research has rich potential for advancing our understanding of theoretical relationships among factors that contribute to school improvement as well as informing policy discourse on alterable factors that contribute to student learning (Creemers and Kyriakides, 2008; Hallinger, 2011; Hattie, 2009; Hawley and Rosenholtz, 1984; Kyriakides *et al.*, 2009).

The current study was framed to contribute to global discourse on the means by which leadership contributes to school improvement. More specifically, the research examines the nature of the paths that link leadership, teaching and learning in schools (Creemers and Kyriakides, 2008; Hallinger and Heck, 1998; Leithwood *et al.*, 2010a, b). The research questions guiding this study were:

RQ1. Do the quality of the school's instructional environment and the effectiveness of its teachers make a difference in student learning in math?

RQ2. What is the relationship among school leadership, effective teaching, and student learning in math?

In order to address these questions, the researchers analyzed a three-year, longitudinal data set constructed from students in the classrooms of several hundred teachers working in 60 primary schools in the USA. The data set combined information drawn from teacher surveys focussing on leadership practices and the instructional environment with value-added estimates of teacher effectiveness, and student math achievement data.

The study's contributions lie in three specific areas. First, the research responds to calls for quantitative longitudinal research on the processes that contribute to school improvement (Heck and Hallinger, 2009, 2010; Luyten *et al.*, 2005; Reynolds *et al.*, 2000; Thoonen *et al.*, 2012). Second, the study demonstrates how a "cross-classified data structure" can be used to examine student achievement results. We highlight this approach in the belief that researchers require a better understanding of the relative merits of different approaches to constructing and analyzing longitudinal data collected from schools. Finally, this research represents an initial empirical effort to model the effects of principal leadership and teacher effectiveness within a single multilevel study. The results further substantiate the relative size of teacher and principal effects on student learning and offer insight into how leadership effects "trickle down" into classrooms and resulting improved student learning.

Theoretical perspective

In this section, we define the theoretical model and major constructs examined in this study. As indicated above, a key contribution of this study lies in its attempt to examine changes in leadership, teaching and learning over time. Therefore, explicit articulation of the study's conceptual model is essential to understanding the relevant relationships among these constructs.

Conceptual model

The conceptual model proposed for this research is represented in Figure 1. The primary constructs are defined as ovals, indicating that they are conceived as underlying processes measured by one or more observed indicators. Context factors (e.g. student demographics, staffing characteristics) are shown as rectangles, indicating that they serve primarily as control variables.

At the school level, the model proposes that school leadership influences student learning outcomes by enhancing the quality of the school's instructional environment.

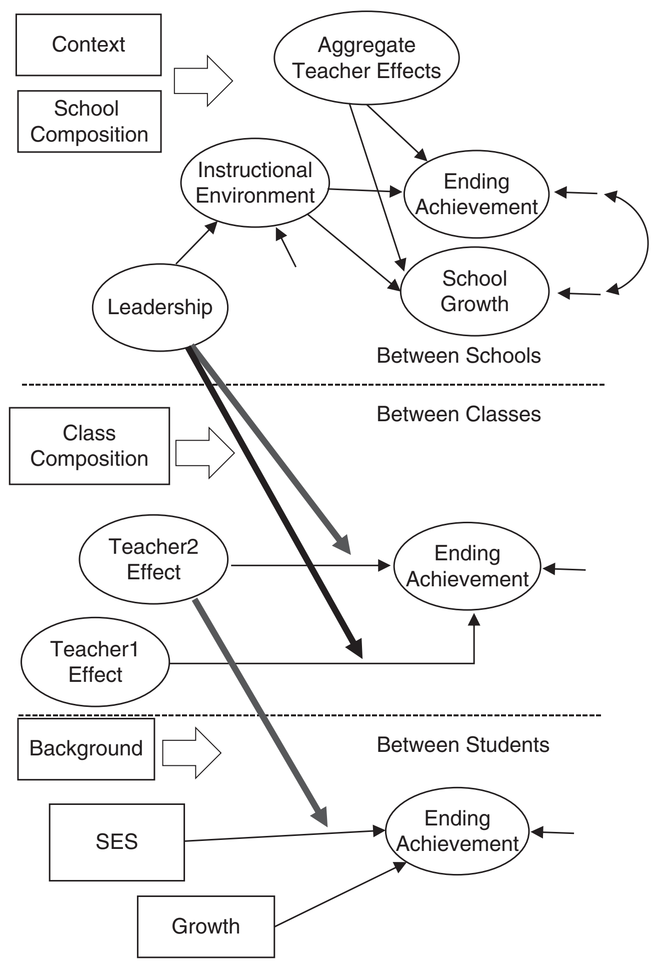


Figure 1.
Proposed conceptual model examining the impact of instructional leadership, instructional practices, and teacher effectiveness on student learning in math

When conceptualized within a dynamic model of school improvement (e.g. Hallinger and Heck, 2010, 2011; Heck and Hallinger, 2009, 2010), leadership effects on student learning become visible in a "downstream" process or "causal chain" that unfolds over time. Here, we propose that the school's instructional environment represents one key path that connects leadership and student learning (see Bossert *et al.*, 1982). This is based on the assumption that teaching and learning represent the "core work" of schools (Bossert *et al.*, 1982; Erickson, 1967; Hallinger and Heck, 1998; Heck and Hallinger, 2009; Leithwood *et al.*, 2004, 2006, 2010a, b; Marks and Printy, 2003; Printy *et al.*, 2009; Robinson, 2011; Robinson *et al.*, 2008).

This kind of conceptualization is often classified as an "indirect effects" or "mediated effects" model of leadership for learning (Hallinger and Heck, 1996, 1998; Leithwood *et al.*, 2004, 2006; Pitner, 1988). In an indirect effects model, the effects of leadership are "mediated" by features of the school (e.g. school climate, capacity, etc.) and its classrooms (e.g. instructional practices). The distinctiveness of the model proposed in Figure 1 lies in the explicit inclusion of information about the school's instructional environment collected from teachers as a mediating construct at the school level. This contrasts with past leadership studies that have typically examined the mediating effects of school conditions such as culture, climate and staff capacity but seldom included classroom-level measures of teaching quality (e.g. Geijsel *et al.*, 1999, 2003; Hallinger *et al.*, 1996; Heck and Hallinger, 2009, 2010; Heck *et al.*, 1990; Ko *et al.*, 2012; Kruger *et al.*, 2007; Leithwood and Jantzi, 2000; Marks and Printy, 2003; Opdenakker and Van Damme, 2007; Wiley, 2001). We expect that principal leadership has a positive effect on these efforts collectively at the school level, as well as on the instructional performance of teachers in their classrooms (i.e. teacher effectiveness).

A second downstream process concerns the role of leadership in coordinating the work of individual teachers. Hence, our model also affords the opportunity to examine the possible "cross-level" impact of school leadership on the contributions of individual teachers to student learning. This proposed multilevel model would enable us to assess and compare the relative contributions of leadership, school level conditions and teacher effectiveness to change in student learning more directly.

We note that the relationships proposed in the causal chain depicted in Figure 1 can be specified as either mediated or moderated (Baron and Kenny, 1986; Preacher, 2011)[1]. A mediated effect implies that school leadership affects classroom achievement indirectly through its impact on individual teacher quality and effectiveness. Conceptualizing leadership as having a moderated effect implies that individual teacher effects on student learning are enhanced in school settings where school leadership and classroom instructional environments are perceived in more positive terms and vice versa. This is consistent with the view that groups partially define individual behavior.

These moderating effects are represented by the vertical arrows in Figure 1 which extend from school leadership to the arrows representing the average impact of each consecutive classroom teacher on student learning (or to the slope coefficients in Figure 3). This model allows us to compare the relative contributions of school leadership and classroom instruction on changes in student learning in greater detail. We suggest that this type of theorizing and comprehensive model testing represents an important incremental step in building theory in school improvement research (see Creemers and Kyriakides, 2008; Heck and Hallinger, 2009; Leithwood *et al.*, 2010a, b; Reynolds *et al.*, 2000; Teddlie and Stringfield, 1993).

Previous school effects research has identified teacher variables as central to the production of student learning outcomes (Creemers and Kyriakides, 2008; Hawley and Rosenholtz, 1984; Kyriakides *et al.*, 2009). Research dating back to the 1970s has highlighted variation among teachers in their "production" of student outcomes (e.g. Hanushek, 1971; Hawley and Rosenholtz, 1984; Murnane, 1975; Rosenshine, 1983). Recent research on teacher quality finds differences in teachers' observable attributes (e.g. experience, educational attainment) and credentials (Goldhaber, 2002; McCaffrey *et al.*, 2004; Sanders and Horn, 1998). Teacher effectiveness research has also advanced our understanding of factors related to teacher classroom behaviors and the achievement of their students (Creemers and Kyriakides, 2008; McCaffrey *et al.*, 2004; Milanowski, 2004; Rosenshine, 1983; Sanders and Horn, 1998). This body of "teacher effectiveness research" has described and measured teaching and learning in various ways including through classroom observations, questionnaires, and value-added estimates derived from student achievement (McCaffrey *et al.*, 2004; Seidel and Shavelson, 2007).

Leadership

The term "leadership for learning" has gained recent international currency (e.g. Hallinger, 2011). In our view, this term draws upon two discrete conceptualizations of school improvement leadership, instructional leadership and transformational leadership (Geijsel *et al.*, 1999, 2003; Hallinger, 2003; Kruger *et al.*, 2007; Leithwood *et al.*, 2004, 2006, 2010a, b). Leadership for learning incorporates the teaching and learning focus of instructional leadership as well as the distinct and more general capacity-building perspective of transformational leadership (Hallinger, 2003; Leithwood *et al.*, 2006, 2010a, b; Marks and Printy, 2003; Mulford and Silins, 2009; Printy *et al.*, 2009).

Four assumptions frame our approach to investigating the paths through which school leadership contributes to student learning and school improvement. First, the effects of school leadership on students are largely indirect (e.g. Hallinger and Heck, 1996; Kleine-Kracht, 1993; Leithwood and Jantzi, 2000; Robinson *et al.*, 2008; Witziers *et al.*, 2003). Second, leadership in schools tends to be distributed among a variety of people in different roles (Day *et al.*, 2011; Leithwood *et al.*, 2004; Marks and Printy, 2003). Third, leadership creates conditions that build capacity for professional learning and change (Creemers and Kyriakides, 2008; Geijsel *et al.*, 1999, 2003; Kyriakides *et al.*, 2009; Leithwood *et al.*, 2004; Marks and Printy, 2003; Opdenakker and Van Damme, 2007; Robinson *et al.*, 2008; Wiley, 2001). Fourth, leadership that increases the school's capacity for improving teachers' instructional practice will enhance student outcomes (Geijsel *et al.*, 1999, 2003; Heck and Hallinger, 2009, 2010; Leithwood *et al.*, 2004, 2006, 2010a, b; Mulford and Silins, 2003; Robinson *et al.*, 2008; Stoll and Fink, 1996).

Theoretical discourse has suggested that factors at the school level can have both direct and indirect effects on student achievement (Hawley and Rosenholtz, 1984; Hattie, 2009; Kyriakides *et al.*, 2009; Pitner, 1988). For example, in their meta-analysis of school factors affecting student learning, Kyriakides *et al.* (2009) concluded that leadership effects were not directly related to student learning outcomes. Consistent with other recent research (e.g. Leithwood *et al.*, 2004, 2006, 2010a, b; Hallinger and Heck, 1998; Heck and Hallinger, 2009, 2010; Marks and Printy, 2003; Robinson *et al.*, 2008; Sebastian and Allensworth, 2012; Witziers *et al.*, 2003), they argued for the need to focus on the targets or paths of leadership as a means of understanding its contribution to school improvement. Consequently, as portrayed in Figure 1, at the school level, we propose that the effects of leadership on student learning will be entirely indirect. This kind of relationship is known as full mediation (MacKinnon, 2008; Preacher, 2011).

This specification contrasts with partial mediation, in which the independent variable (leadership) is also proposed to exert a statistically significant direct effect on the dependent variable (Preacher, 2011).

School-level instructional environment

Consistent with the notion of a fully mediated model of school leadership effects (e.g. Heck and Hallinger, 2010; Leithwood *et al.*, 2010a, b), this study focusses on understanding the relationship between leadership and teaching-related variables. The first of these teaching-related variables is represented by the school's instructional environment. We note that although leadership is conceived to operate at the school level, the instructional practices of teachers are enacted in an aggregate of individual classrooms. More specifically, as Bidwell (1965) observed, the organization of schooling provides teachers with professional discretion to make individual judgments regarding students' needs and abilities. This is necessary so that they can make needed adjustments in day-to-day instruction al activities. At the same time, however, the necessity for teachers to deliver a formal curriculum requires considerable uniformity and routinization in moving students sequentially from grade to grade and school to school within the education system.

Networks of informal faculty relationships serve both to buffer and further stabilize formal classroom and school structures, while situating important mechanisms through which teachers can adapt their work to external demands (Bidwell, 2001). Balancing the needs for teacher autonomy and systemic uniformity, while building professional relationships, therefore, represents a primary task of school leaders. We suggest conceptual models that seek to describe interactions between leadership and classroom teaching must take into account the fact that these "processes" are operating at different organizational levels. The conceptual model must, therefore, clearly specify these school-to-teacher relationships, as is shown in Figure 1.

This construct describes policies and practices that create conditions which impact the quality of classroom instruction. These include, for example, the quantity of time spent on teaching, student opportunity to learn, extent and nature of teacher collaboration, and teacher participation in professional development. We examine school policies and actions that directly support classroom teaching (e.g. teachers' use of time and preferred instructional strategies), develop the school's learning environment (i.e. teacher collaborative efforts to develop and refine the curriculum), enable the professional learning of the staff, and provide a collaborative means for implementing strategic actions aimed at continuous improvement (e.g. Creemers and Kyriakides, 2008; Leithwood *et al.*, 2004; Mulford and Silins, 2009; Opdenakker and Van Damme, 2007; Stoll and Fink, 1996). As we detail in the methods section, we gathered individual teachers' perceptions of the quality of the instructional environment in their classrooms and schools. Then we used multilevel factor analysis to derive a school-level measure of the quality of the learning environment within each school.

Classroom-level teacher effectiveness

Figure 1 further suggests that leadership can influence student learning by creating conditions that enable more effective teaching among individual teachers, or what we refer to as teacher effectiveness. We note that testing this hypothesis is not straightforward due to the increased complexity involved in accounting for different combinations of students moving through classrooms with teachers' of different skills over time. Indeed, studies of teacher effectiveness often fail to meet the full requirements for modeling these relationships, since they do not include school-level processes (e.g. Sanders and Horn, 1998). These school-level actions and organizational routines provide an organizational milieu in which individual teachers' instructional practice is to a greater or lesser extent coordinated (Bidwell, 1965).

Growth in student learning

Assessing growth in the achievement of students over time presents a variety of conceptual and technical challenges (Hallinger *et al.*, 2014; Murphy *et al.*, 2013). For example, when seeking to understand "school improvement in learning" researchers can represent growth in student achievement by tracking growth of student cohorts or of individual students. Longitudinal assessment of individual student progress through school is superior to comparing successive student cohorts (e.g. percentages of students who attain proficiency each year) for the purpose of monitoring school improvement (Willms, 1992). Monitoring the progress of individual students over time captures their actual academic growth as they move through their educational careers. This focusses attention more squarely on students' experiences in attending a particular school over multiple years and with multiple teachers. Moreover, it provides a direct means of recognizing that schools serve students who start at different places and progress at different rates (Seltzer *et al.*, 2003).

As summarized in Figure 1, the key premises of our study are that the quality of the school's instructional environment, as perceived by individual teachers, mediates the effect of school leadership on school outcomes and, simultaneously, school leadership moderates the estimated effects of individual teachers on student learning within their classes. Moreover, we suspect that the sum of individual teacher effects on student learning contribute to the collective school milieu of instructional effects on school outcomes. This model provides an initial specification of how multilevel strategic leadership action can enhance school improvement.

Method

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, etc. 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. (Hallinger and Heck, 2011, p. 154)

The relevance of employing longitudinal analysis in studies of school improvement has recently been brought into sharper focus as school effectiveness researchers have increasingly accepted the conceptual argument behind studying school improvement as a time-mediated process (Hallinger and Heck, 2011; Heck and Hallinger, 2010; Mulford and Silins, 2009; Thoonen *et al.*, 2012). The adoption of longitudinal research designs challenges researchers to employ new methods for data collection and analysis.

There are, for example, a number of different ways to incorporate a temporal element into school improvement studies. One common approach for conceptualizing and organizing the collection of longitudinal data is through collecting repeated measures on individuals' outcomes and explanatory variables. For example, Lazarsfeld and Fiske (1938) initially proposed the cross-lagged approach as an alternative to cross-sectional data modeling when seeking to assess the direction of effects between measured variables in a proposed model.

Cross-lagged longitudinal models imply that two (or more) variables may be both a cause and an effect of each other over time (Kline, 2004; Marsh and 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 subsequent intervals (e.g. Times 2, 3, 4, etc.) or at some specified time "lag" (Hallinger and Heck, 2011, p. 156; Oud, 2002). A typical classroom example might be that teachers respond to students' low performance in math by altering instruction to re-teach and practice a particular needed skill, which increases test performance, which may again alter future instructional strategies.

Growth curve models are a related type of longitudinal model. In this approach, change in the achievement of individuals over time may be linked to both time-invariant and time-varying covariates (Raudenbush and Bryk, 2002). Growth curve models are useful in examining students' change within a particular teacher's class from beginning to end of the year through repeated measurements, or in assessing student change over several grade levels within a particular a particular school.

Cross-lagged and growth curve models may not, however, always suit the requirements of a particular research situation. For example, when studying school improvement over multiple years, researchers may wish to account not only for changes in the performance of students within schools and between schools, but they may also wish to link those changes to the different combinations of teachers to whom the students were assigned (e.g. see Borman and Kimball, 2005; Milanowski, 2004; Sanders and Horn, 1998). This requires a more complex nesting of students within different combinations of teachers as an additional hierarchical level within the data set (see also Hallinger and Heck, 2011; Hallinger *et al.*, 2014).

In advancing leadership research, this data arrangement becomes essential if we hope to illuminate how leadership effects on learning are mediated by changes in the practices of different teachers. Such research may necessitate dealing with "cross-classified" data (e.g. students having a different teacher each year) or perhaps even "multiple membership" data (i.e. students having multiple teachers each year). The first arrangement is common in primary schools in the USA, while the second is common in secondary schools in the USA.

In our study, we examine this first type of cross-classified hierarchical data structure common in primary schools. This actually necessitates a four-level data structure. At level 1, repeated measures in math are nested within students over two years. At level 2, we can examine differences between students in achievement. At level 3, students are cross-classified within two successive primary classrooms. This means that any two students in the data base could have common first- and second-year teachers, one common teacher (i.e. either the first or second teacher), or no teacher in common. At level 4 we define the school context and process variables.

To help articulate this multilevel structure, in Figure 2 we illustrate the cross-classification of students within teachers at level 3 for the first ten students in our sample. The figure shows that these ten students had two different fourth grade teachers located in two different schools (schools 1 and 2). Seven students had the first teacher and three had the second teacher. In the fifth grade, they went to four different teachers. Students 1-3 went to teacher 1 in school 1, students 4-7 went to teacher 2 in school 1, students 8-9 went to teacher 3 in school 2. Finally, student 10 (from fourth grade teacher 2) went to teacher 4 in a third school.

As Figure 2 implies, cross-classified models take into account the fact that during their primary years, students study with different combinations of teachers.

Moreover, these combinations may have differential consequences for their learning outcomes due in part to the differential quality of teachers and peer arrangements (McCaffrey *et al.*, 2004). Cross-classification of students makes the teacher-level data set more complex than the typical multilevel, cross-sectional study or the repeated measures study where students are only nested within one specific classroom and school. Indeed, as the study duration lengthens, more and more different teacher-student combinations enter into computations. Thus, it becomes increasingly challenging to estimate the effects of each individual teacher on student outcomes, since longer time frames introduce problems relative to accounting for diminishing teacher effects; for example, the year-1 teacher's effect on student outcomes at year 5 (McCaffrey *et al.*, 2004). Moreover, it can require having outcome measures that are vertically equated over several years to permit repeated observations of students' progress on one integrated assessment instrument.

In our study, these problems are minimized, since we only examine the effects of two successive teachers on student outcomes. Of course, limiting the duration of the study to approximately three years simultaneously limits our ability to detect patterns of change fully, since school improvement processes may require three to five years to become fully apparent (Hallinger and Heck, 2011).

Sample

The final sample used in this study consisted of 2,894 students cross-classified in 240 fourth and 163 fifth grade classrooms in 60 primary schools in one western state in the USA. These data represent a subset of a larger data set examining teacher effectiveness (Heck, 2009). The data set combined information drawn from teacher surveys about their schools' leadership practices and classroom instructional environments in year 1. We developed value-added teacher effectiveness scores from a previous student cohort during years 1 and 2, and resulting math achievement data of the current student cohort in their classrooms during years 2 and 3.

The data set is unique in several ways in terms of informing about school improvement. First, it allows the development of a clear set of temporal relationships. With respect to teachers, for example, it allows us to develop value-added estimates of teacher effects on an earlier student cohort (i.e. during years 1 and 2) and then apply those scores subsequently to determine whether they impact a second cohort, such that teachers' effectiveness derived scores are independent from the current students in the study of school improvement. This provides one test of the validity of the prior teacher estimates in differentiating student achievement at a subsequent interval. Second, it facilitates the estimation of leadership effects on later achievement of students

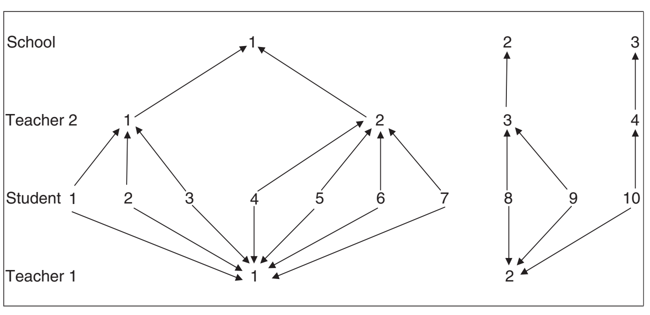


Figure 2.
Illustration of cross-classified data structure for the first ten students in the study

(and later moderating effects on teacher-student effects) through prior information collected from individual teachers about their school's leadership practices and its instructional environment before the student background, teacher effectiveness, and achievement data were collected. Third, the study offered an opportunity to examine cross-level relationships between school- and classroom-level structural arrangements and corresponding environments that have been proposed to impact growth in student learning, but have not been considered in one simultaneous multilevel model. In this way, our analyses seek to identify the paths that link school leadership, teacher instructional practices, and growth in student learning during the process of school improvement.

Variables in the models

Context variables. These included a range of student background variables (gender, low SES (as identified by participation in the federal free/reduced lunch program), underrepresented group by race/ethnicity in postsecondary education, language background), classroom context (student composition), and school context variables (enrolment size, student SES, special education, and language composition, staff stability). We note that in a cross-classified model, it is redundant to enter composition data for each of the classrooms (McCaffrey *et al.*, 2004). Because of our focus on ending student achievement as well as growth over the two-year period, we also entered classroom composition data with respect to the second classroom.

School leadership. The leadership construct was defined by ten items which are consistent with our four preliminary assumptions about collaborative leadership. These include shared vision with clear goals and collaborative work, shared among role groups in the school, focussed on building conditions for professional learning and change, and directed toward improvement of conditions that support student learning. The items measured three specific aspects of shared, or collaborative, leadership (with items paraphrased in parentheses):

- Shared focus on instruction for student learning (e.g. student achievement is top school priority; school has clear academic mission and goals; administrators and staff work toward identifying and resolving instructional problems; effective two-way communication exists; school administrators regularly observe and provide feedback to teachers regarding their instruction; school administrators provide feedback to teachers regarding instruction).
- Collaborative decisions focussing on academic improvement (e.g. school leaders and teachers work cooperatively to develop a school improvement plan; administrators and teachers share in leadership roles; teachers have the needed instructional resources to teach effectively).
- Focussed participation in efforts to evaluate the school's academic development (e.g. school has an effective, ongoing system for evaluating progress toward reaching academic goals).

Table I provides the results of our two-level confirmatory factor analysis (CFA) from the teacher survey data (i.e. 4,056 teachers within the 60 schools) to develop a weighted factor score comprised of the ten indicators of school leadership practices in their school[2]. Items used to describe leadership were measured on five-point, Likert-type scales. At the teacher level, the table suggests the item loadings were all statistically significant and ranged from 0.35 to 1.00. Cronbach's α ($\alpha = 0.85$), a measure of subject

	Estimate	SE	Two-tailed Est./SE	*p*-value
Within schools (n = 4,056 teachers)				
Leadership				
Achievement is top goal	1.000	0.000	******	0.000
Clear mission and goals	0.624	0.020	31.276	0.000
Works to resolve				
Instructional problems	0.574	0.017	33.375	0.000
Observes instruction	0.495	0.014	34.291	0.000
Provides feedback	0.348	0.017	19.893	0.000
Collaborates on school				
Improvement plan	0.483	0.018	26.512	0.000
Shares leadership	0.360	0.018	19.651	0.000
Sufficient resources	0.475	0.018	25.824	0.000
Ongoing evaluation	0.451	0.020	22.428	0.000
Instructional environment				
Time on instruction	1.000	0.000	******	0.000
Maximize time on task	0.667	0.013	51.017	0.000
Active learning	0.777	0.013	62.010	0.000
Present in varied ways	0.228	0.019	11.820	0.000
Use varied strategies	0.656	0.013	49.266	0.000
Collaborate on curriculum	0.692	0.014	51.035	0.000
Professional develop	0.695	0.013	53.788	0.000
Refine teaching	0.740	0.015	49.989	0.000
Extra help provided	0.699	0.015	47.800	0.000
Time to master skills	0.773	0.012	66.887	0.000
Between schools (n = 60 schools)				
Leadership				
Achievement is top goal	1.000	0.000	********	0.000
Clear mission and goals	0.985	0.001	1237.436	0.000
Works to resolve				
instructional problems	0.983	0.001	1863.458	0.000
Observes instruction	0.961	0.001	875.151	0.000
Provides feedback	0.952	0.001	761.486	0.000
Collaborates on school				
improvement plan	0.981	0.001	1207.534	0.000
Shares leadership	0.975	0.001	802.949	0.000
Sufficient resources	0.988	0.001	1707.338	0.000
Ongoing evaluation	0.974	0.001	1423.901	0.000
Instructional environment				
Time on instruction	1.000	0.000	********	0.000
Maximize time on task	0.994	0.001	1129.050	0.000
Active learning	0.988	0.000	3920.413	0.000
Present in varied ways	0.991	0.003	345.545	0.000
Diverse varied strategies	0.984	0.001	1331.255	0.000
Collaborate on curriculum	0.972	0.001	750.383	0.000
Professional develop	0.964	0.000	2352.653	0.000
Refine teaching	0.974	0.002	632.970	0.000
Extra help provided	0.973	0.001	1121.107	0.000
Time to master skills	0.983	0.000	3743.050	0.000

Note: Astericks indicates the factor loading is fixed to 1.0 to establish a metric to measure the latent factor

Table I.
Between- and within-school CFA of items comprising constructs standardized estimates

internal consistency, provides another indicator of the reliability of this construct. At the school level, the item loadings on the leadership factor were all above 0.95 (see Table I). The between-school leadership proportion of the factor scores were then saved in the data set and used in subsequent multilevel analyses.

We note that one of the advantages of using multilevel CFA is to produce collective scores about school leadership and instructional practices that have been corrected for the unreliability of individual teachers' perceptions as represented through the separate items comprising the factors (Muthén, 1994). This does not happen when scores are simply summed across individuals within each school and the mean is used to represent the school estimate or where the factor analysis is simply done at the teacher level and the scores aggregated in a similar manner.

School-level instructional environment. Teacher survey data were also used to develop a weighted factor score comprised of ten indicators of classroom and school practices. These reflected our definition of the school's instructional environment (e.g. support for students with special learning needs, collaboration among teachers; participation in professional development). Specific items included: teachers use class time primarily for instruction; teachers maximize students' instructional time on task; teachers present academic work in varied ways; teachers use varied instructional strategies to help all students learn; instruction is geared to active student learning; teachers collaborate to develop and refine academic curriculum; teachers participate in professional development activities regarding instructional practices; teachers continually assess their instruction to refine teaching; students who need extra help receive it; students are given adequate time to master skills.

At the teacher level, the results of our analysis presented in Table I suggest that the factor loadings for the items were all statistically significant (ranging from 0.23 to 1.00). The corresponding Cronbach's α for individual teachers was also strong at 0.91. The school-level items all loaded above 0.95, again indicating very little school-level error after correcting the individual teacher scores for error. Between-school factor scores were then saved and used in the subsequent multilevel analyses. Overall, the results in Table I provide evidence that the two key constructs our study were measured quite well at the school level.

Teacher effectiveness. Recently, researchers infer teacher effectiveness from a residual of the teacher's current (or previous) class achievement obtained through statistical adjustment (Kupermintz, 2003; McCaffrey *et al.*, 2004). A residual, defined as the observed score minus an expected score, is an adjusted estimate of the teacher's effect after removing the effects of other extraneous variables (i.e. student background, previous achievement). Such residual, or "value-added," estimates can be a useful way of describing teacher effectiveness, but they depend on the validity of the statistical models from which they are derived (McCaffrey *et al.*, 2004). It is often helpful to include classroom and school contextual variables in order to aid in the interpretation of the relative size of the teacher effect on student outcomes (Kupermintz, 2003; McCaffrey *et al.*, 2004).

In our study, the teacher effectiveness scores primarily reflect student learning outcomes resulting from being assigned to a particular teacher, rather than the teacher's actual behavior to produce the outcomes. The estimates of teachers' effectiveness in producing student achievement for classrooms in years 2 and 3 of the study were generated from the background and achievement of the previous student cohort assigned to the teachers (i.e. during years 1 and 2) and we estimated from a larger data base of over 1,000 teachers in 160 primary schools (Heck, 2009).

The use of the subset of 60 schools resulted from having the item-level data from the teachers in these schools to develop the leadership and instructional environment constructs. We relied on teacher data in this study because they are closest to these key school processes – that is, teachers are primarily responsible for their individual classroom practices, and they each have a separate (and to some extent unique) relationship with the school principal. We concentrate on separating the unique parts from the collective part within each school through our preliminary CFA.

In estimating the teacher effectiveness scores, we also caution it can be difficult to disentangle the various internal contexts, structures, and processes (e.g. higher academic expectations, more advanced curriculum) from individual teacher effects (McCaffrey *et al.*, 2004). Because of the correlation between aspects of instruction and class context, adjustments for contexts can underestimate "true" teacher effects (De Fraine *et al.*, 2002). Student composition and peer relationships can also influence various educational processes within each school including teacher academic expectations, grouping processes, curriculum coverage between classrooms within grade levels, and the social distribution of learning (Lee and Bryk, 1989). We acknowledge the presence of these factors but, in general, we do not hypothesize about their specific effects (i.e. as suggested by the broad arrows included in Figure 1). One exception is we propose that stronger teacher effectiveness should diminish the social distribution of learning (e.g. as indicated by socioeconomic status) within classrooms (i.e. implied by the vertical arrow in Figure 1 from teacher 2 to the slope describing the effect of student socioeconomic status on achievement).

At the school level, we aggregated the individual teachers' effectiveness scores and examined the relative level of these schools and their relative variability in explaining differences in math achievement between schools. We define the level of average teacher effectiveness as the mean of individual teachers' effectiveness over the two-year period of our study. In addition, we examine the potential variability in average teacher effectiveness within each school in contributing to student achievement. We hypothesize that both the average level of teachers' effectiveness, as well as the variability in their effectiveness will also contribute to student learning beyond individual teacher effects. More specifically, we propose that less variability in effectiveness among teachers will be related to stronger school-level achievement. In our model, therefore, the individual effects of teachers as well as their aggregated effects may accumulate to provide some academic advantages to students.

Growth in math achievement. The achievement outcome in this study was student scaled scores on the Stanford Achievement Test (SAT) version 9 measured over two years. The scores are vertically equated, which permits comparison of individuals' scores over multiple occasions. For interested readers, we provide the basic specification[3] for model 1 (see Table II).

As shown in Figure 1, we capture growth student learning, our measures of school improvement, in two parameters. At the student level (level 1), we define the achievement intercept to indicate students' level of math achievement at the end of the study (i.e. fifth grade achievement) rather than their initial status achievement measured at the beginning. This specification is different from how most growth models are formulated. Therefore, in examining schools' improvement we focus on explaining where students end up instead of where they begin. We define student growth in math as the linear change in math scores of the current student cohort during the two-year period. In this type of model specification, where students "end up" is typically positively related to how much they gain between the repeated measures.

Variables	Mean	SD
Student level (n = 2,794)		
Ending math	659.09	44.03
Female	0.50	na
Minority	0.42	na
Low SES	0.38	na
Other language	0.20	na
Classroom 1 (n = 240)		
Teacher 1 effectiveness	−0.06	0.86
Classroom 2 (n = 161)		
Teacher 2 effectiveness	0.07	0.70
Student composition	0.41	0.27
School level (n = 60)		
Average teacher effectiveness	0.07	1.00
Average variability in effectiveness	0.76	0.64
Staff stability	0.71	0.17
Enrollment	469.84	146.69
Student composition	0.05	1.19
Leadership	0.00	1.00
Instructional practice	0.00	1.00

Note: na, not applicable (no standard deviation)

Table II. Descriptive statistics

Results

We present the results of the study in terms of the several steps involved in testing the conceptual model proposed in Figure 1. We then interpret the findings with respect to the specific research questions in the concluding section of the paper.

Descriptive statistics

Descriptive statistics are presented below in Table II. Student background variables indicate a relatively high proportion of students (0.38) participating in the federally free/reduced lunch program (defined as low student socioeconomic status). In addition, a considerable number of students spoke languages other than English at home (0.20), and a considerable number of students was identified as underrepresented by race/ethnicity (0.42) with respect to participating in postsecondary education. We note that at the classroom level, the mean of the student composition variable ($M = 0.42$) suggests the average class in this study consisted of relatively more students of low SES background and who spoke languages other than English at home than the typical classroom in the larger data base (i.e. about 0.42 SD above the grand mean).

We also note that year 2 teachers were slightly more effective on average ($M = 0.07$) compared with year 1 teachers ($M = -0.06$). There is, however, considerable variability in both teachers 1 and 2 effects (SD = 0.86, 0.70, respectively). In contrast to classrooms, school student composition was relatively normally distributed among participating schools ($M = 0.05$, SD = 1.19). The average school enrolled approximately 470 students (SD = 147) and had an average staff stability mean of 0.71 (i.e. defined as the proportion of teachers in the school for past five years). As we noted previously, the leadership and instructional composition means ($M = 0$) reflect latent factor means at the school level corrected for individual errors of measurement. Finally, the mean and standard deviation for variability in teacher effectiveness ($M = 0.76$, SD = 0.64) reflect considerable differences across schools in how tightly configured or widely dispersed classroom teachers were in terms of their levels of effectiveness.

Model 1: estimating ending math achievement and growth

Model1 1 (see Table III) provides an initial benchmark for estimating the average ending achievement and growth in math for schools in the study. The average ending achievement for schools was 656.718; average growth over the period was 29.022. This implies that students grew about 29.02 scaled score points in math between fourth and fifth grade, or about 0.55 of a standard deviation (not tabled). We note that this yearly gain is consistent with Bloom *et al.* (2008), who found average annual gains in math were 0.56 SD between grades 4 and 5 on comparisons of seven standardized tests (including the SAT9 used in this study).

Variable	Model 1 estimate	Model 2 estimate	Model 3 estimate	Model 3 std. estimate
Intercept model	656.718*	683.399*	692.644*	0.000
School composition		−8.273*	−8.095*	−0.154*
Enrollment		2.986	3.523	0.067
Teacher stability		0.819	0.693	0.013
Instructional environment			14.690*	0.280*
Leadership		5.455	0.244	0.005
Average teacher effectiveness		0.530	0.456	0.009
Variability in teacher effectiveness		−4.266*	−3.991*	−0.076*
Teacher 2 effectiveness		11.189*	11.144*	0.213*
Leadership × Teacher 2		6.553*	6.523*	0.124*
Teacher 1 effectiveness		10.877*	11.103*	0.212*
Leadership × Teacher 1		2.647	2.767	0.053
Class composition		−0.106	−0.085	−0.002
Female		2.203*	2.218*	0.042*
Minority by race/ethnicity		−15.101*	−15.147*	−0.289*
Other language		−0.105	−0.084	−0.002
Low SES		−12.765*	−12.754*	−0.243*
Low SES × Teacher 2 effectiveness		2.693**	2.686**	0.051**
Growth model	29.022*	29.321*	33.317*	0.635*
School composition		−1.086*	−1.013*	−0.019*
Enrollment		1.109	1.397	0.027
Teacher stability		−0.931	−0.962	−0.018
Average teacher effectiveness		2.729*	2.681*	0.051*
Instructional environment			6.265*	0.119*
Leadership		−0.455	−2.707	−0.052
Standardized variance components				
School variance	0.094	0.013	0.011	0.011
Teacher variance	0.060	0.033	0.031	0.031
Individual variance	0.523	0.501	0.501	0.501
School growth variance	0.004	0.001	0.001	0.001
Deviance	12,379.732	11,982.085	11,975.448	11,975.448
Parameters	7	28	30	30

Notes: $*p < 0.05$; $**p < 0.10$

Table III.
Comparing models

The variance components in Model 1 can also be used to estimate the proportion of total variance in math achievement attributable to students, classrooms, and schools (i.e. estimated as the variance for each level divided by the total variance). The results suggest that about 14 percent of the variance in math achievement was between schools ($0.094/0.677 = 0.14$), with about 9 percent due to the cross-classification of students within teachers ($0.060/0.677 = 0.09$). This latter estimate suggests that, for any two students, having the same two teachers in common accounted for about 9 percent of their ending math scores. This variance was about evenly split between first and second year teachers (not tabled). Finally, about 77 percent of the variance was due to student differences ($0.523/0.677 = 0.77$).

Models 2 and 3: testing-mediated leadership effects

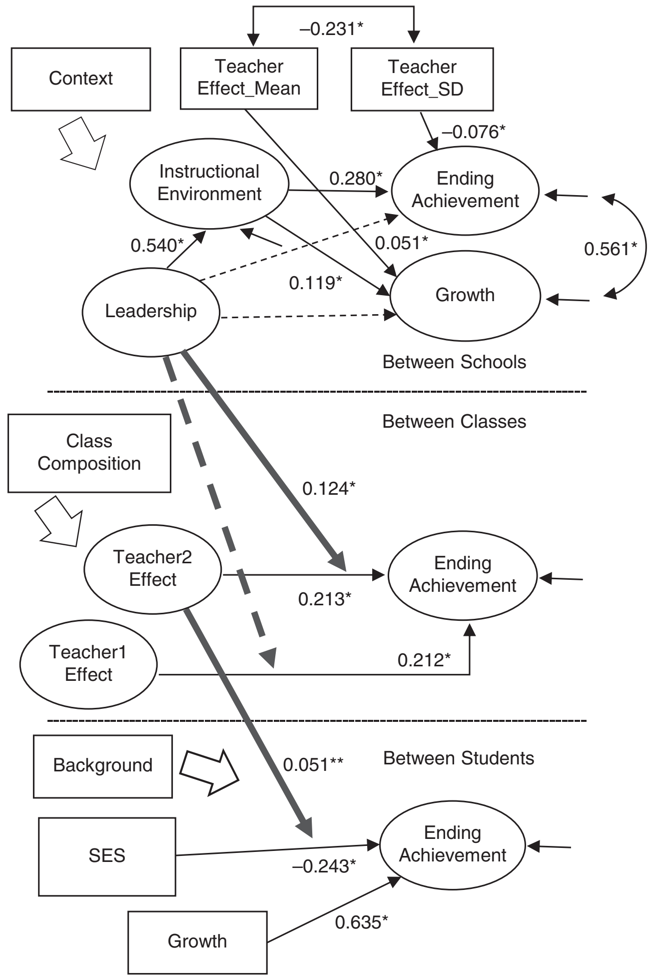
Models 2 and 3 in Table III provide the estimated direct (and indirect) effects attributed to school leadership in explaining ending school math achievement and growth. When testing a proposed mediated effect, the first step is typically to propose a model with only the direct effect of X to Y specified (MacKinnon, 2008). Model 2, therefore, summarizes the direct effect of leadership on ending achievement and growth, without the presence of a mediator (e.g. instructional environment). A closer look at this model indicates that leadership does not directly affect ending math achievement ($\gamma = 5.455$, $p > 0.10$) or growth in math ($\gamma = -0.455$, $p > 0.10$), after controlling for other variables in the model. Model 2 can be interpreted as a "total effect" model, since it represents the total effect of X (leadership) on Y (student learning), without consideration of possible mediating variables.

Model 3 provides a test of the possible direct and indirect effects of leadership on ending math achievement and growth. There are two possible results of interest when testing for mediated effects. The effect of leadership on school outcomes can be both direct and indirect (i.e. through its relationship with the mediator). This relationship is referred to as partial mediation ($X \rightarrow Y$ and $X \rightarrow M \rightarrow Y$). Alternatively, the effect of leadership may be entirely indirect ($X \rightarrow M \rightarrow Y$), in which case the relationship is referred to as full mediated (Baron and Kenny, 1986; Preacher, 2011).

Model 3, therefore, re-estimates the effects of leadership on ending math achievement and growth in math by including the direct effects of the instructional environment on the outcomes. Testing this kind of model also necessitates estimating a separate path from leadership to instructional environment; this was estimated at 0.357 ($p < 0.05$, not tabled). The effect of the school's instructional environment on ending math achievement was statistically significant ($\gamma = 14.690$, $p < 0.05$). This analysis also provides an adjusted estimate of leadership's direct effect on ending math achievement, which was again not statistically significant ($\gamma = 0.244$, $p > 0.10$).

The two paths are then multiplied to estimate the indirect effects of leadership on math outcomes (MacKinnon, 2008). For ending achievement, our empirical test indicated that the indirect effect of leadership was statistically significant[4] ($\gamma = 5.244$, $p < 0.05$, not tabled). The size of the indirect effect is estimated at 0.357×14.690. In Figure 3, the corresponding standardized indirect effect was similarly estimated as 0.15, suggesting a sizable indirect relationship. Since the direct effects of leadership were not statistically significant, we conclude the effects of leadership on school outcomes in math in this model were fully mediated by the school's instructional environment.

For growth in math, Table III suggests the estimated standardized direct effect of the school's instructional environment on growth was statistically significant



Notes: **$p < 0.10$; *$p < 0.05$ - - - - ▸ nonsignificant, tested

Figure 3.
Standardized effects of final cross-classified model explaining outcomes

(0.119, $p < 0.05$). As expected, the direct standardized effect of leadership on student growth was not statistically significant (–0.052, $p > 0.10$). The standardized indirect effect of leadership on student growth (through instructional environment) was estimated as positive ($0.54 \times 0.12 = 0.065$, see Figure 3); however, the total effect was negated by the negative standardized direct effect (–0.052). There is likely an indirect mechanism that affects student growth, but it likely takes more time for such leadership effects to become observable, given that growth in this study only covers a one-year interval between the first and second teacher.

Examining the moderating leadership effects on effectiveness-learning slopes

We next turn our attention to possible moderating effects in Model 3. Within schools, we assume that stronger perceptions of leadership practices will enhance the effects of individual teachers on student learning. This type of relationship is summarized in

Figure 1 with vertical arrows extending from leadership to the slope relationships between first and second teacher effectiveness on ending math achievement.

First, at the classroom level, the Model 3 results suggest that both teachers contribute meaningfully to ending math achievement, controlling for school-, classroom-, and student-level demographics. We can interpret the combined teacher results as follows. Having two consecutive teachers one standard deviation above the mean in effectiveness would result in a 22.247 point increase in ending math achievement (11.103 + 11.144), compared with students who had successive teachers of average effectiveness, controlling for other variables in the model. After standardizing the scores (see Table II), this translates to about a 0.425 SD increase (0.212 + 0.213, respectively) in ending math achievement.

Second, in Model 3 we observed that leadership had a significant moderating effect on the size of the second teacher's effect on ending math achievement ($\gamma = 6.523$, $p < 0.05$), controlling for other variables. The standardized effect was estimated as 0.124. This suggests that a 1-SD increase in the perceived quality of school leadership would increase the teacher effect by an additional 0.124 SD. Although the cross-level leadership effect was also positive for the first teacher ($\gamma = 2.767$), it was not statistically significant ($p > 0.10$)[5].

Third, several other effects were also of interest in Model 3. We found some support for the preliminary hypothesis that stronger teacher effectiveness would reduce the student-level gap in achievement associated with student SES within classrooms. In terms of standardized effects (as summarized in Figure 3), for a 1-SD increase in teaching effectiveness, we would expect a reduction in the standardized effect on ending achievement due to low SES (−0.243) of about 0.051 (or a reduction of about 21 percent), controlling for other variables. At the school level, Figure 3 suggests lower variability in average teacher effectiveness was related to higher ending math achievement ($\gamma = -0.076$, $p < 0.05$), controlling for other variables. Higher average teaching effectiveness had a small positive standardized effect on growth in student learning (0.051, $p < 0.05$), controlling for other variables. The inverse correlation between them ($r = -0.231$) suggests higher levels of effectiveness go with a tendency toward less variability. Finally, Figure 3 suggests a substantial positive correlation between ending achievement and growth ($r = 0.561$). This supports the view that where students ended in math achievement was substantially related to their growth.

Discussion

Over the past decade, education reforms implemented globally have increasingly evidenced the expectation that school leaders should be capable of improving the quality of teaching and learning in schools (Hallinger, 2011). In some countries, this expectation has become firmly embedded in the fabric of accountability policies that mandate penalties for school principals who "fail to make the grade" (De Fraine *et al.*, 2002; Leithwood, 2001; Nettles and Herrington, 2007; Silva *et al.*, 2011; Walker and Ko, 2011). Thus, establishing an empirical link that validates the nature of the relationship between school leadership and student learning is of importance to policymakers and practitioners as well as researchers.

In this paper, we employed longitudinal data gathered from 60 primary schools in the USA in order to examine school leadership effects on teaching and learning. We tested a series of multilevel, cross-classified models that sought to illuminate how leadership and teaching effectiveness contribute to growth in student learning over time. In this concluding section of the paper, we revisit limitations of the study and then offer our interpretation of key findings.

Limitations of the study

One aim of the current study was to demonstrate the utility of analyzing longitudinal data in studies of school improvement. Despite some success in this effort, we acknowledge that the three-year duration of the study remains an overly constrained period of time in which to illuminate patterns of school change and improvement. Indeed, in another study where we analyzed five years of data, it was possible to identify more substantial patterns of change in leadership and other school improvement processes (see Heck and Hallinger, 2009, 2010).

Even while acknowledging this limitation, we also wish to highlight the fact that it is considerably more challenging to study the kind of "cross-level relationships" specified in the current study over longer periods of time[6]. This is due to various technical problems (e.g. monitoring the progress of students cross-classified with a larger numbers of teachers, having tests that have been equated over longer periods of time, and dealing with the diminishing effects of earlier teachers on later outcomes) as well as practical problems that may arise such as gaining access to stable longitudinal, multilevel data sets[7].

A second limitation of this study concerned the use of a survey to assess teachers' perceptions of their instructional practices. Although the surveys met acceptable standards of reliability, lower inference measures of teacher instructional behavior (e.g. lesson logs, classroom observations) are preferable for developing valid assessments of instructional practices (Good and Brophy, 1986).

A third limitation lies in the use of value-added measures of teacher effectiveness (see Hallinger *et al.*, 2014; Murphy *et al.*, 2013). These represent proxy measures for information that describes what teachers actually do to influence student learning in their classrooms. Although survey data from individual teachers was used to validate the value-added scores representing teachers' contribution to student outcomes, as already noted, surveys themselves are an imperfect means of assessing teaching quality. Thus, future studies would do well to try to obtain better information on teacher effectiveness.

Summary and interpretation of the findings

Despite these limitations, the findings offer rather clear answers to the research questions posed at the outset of the paper. First, we found that both the instructional environment and teacher effectiveness were positively and directly related to students' ending math achievement. At the school level, the effect size (in standard deviation units) for the school's instructional environment on ending math achievement was significant and moderately substantial (i.e. 0.28). This variable also had the largest impact on school growth (0.12) among the school-level constructs.

Consistent with other studies, we also found that "teacher effects" matter when modeling growth in student achievement (Mendro *et al.*, 1998; Rowan *et al.*, 2002; Sanders and Horn, 1998; Seidel and Shavelson, 2007). The effectiveness levels of successive teachers provided a meaningful explanation for change in student achievement. For example, a student who had two consecutive teachers whose effectiveness was 1-SD above the grand mean enjoyed an increase of 0.43 SD in math achievement, compared with students who had successive teachers of average effectiveness.

Moreover, teacher effects appeared to compound from individual classrooms to the school as a whole. This provides a quite small (0.08), but statistically significant, academic advantage to students who were attending primary schools where teachers were more "uniformly effective." Additionally, students in schools that had higher-than-average effectiveness scores also made relatively more growth (0.05) than students in schools of average teacher effectiveness. These effects are generally missed in teacher effectiveness studies that do not include school-level variables.

The effects of leadership for learning represented the focus of the second research question, and here the answer is also unequivocal. We found that instructionally focussed leadership was indirectly but significantly related to ending math achievement through its positive effect on the instructional environment. For ending achievement, the results suggest that a 1-SD increase in the strength of leadership in a school could yield a commensurate 0.15 increase in ending math scores. We note this indirect effect is consistent with the minimum size of school-level variables that Creemers and Kyriakides (2008) adopted as "substantial" in their meta-analysis of school factors that directly impact student achievement. We did not find an observable indirect of leadership on growth. We note, however, our measure of student growth was limited to only one interval.

Although our measure of school leadership represents teachers' collective perceptions of the principal's (and others') leadership, we emphasize that our measurement approach initially adjusted the school-level constructs for measurement unreliability at the individual teacher level. We do acknowledge, however, that teachers' perception of leadership is not the actual leadership itself nor is it exactly the same as principals' perceptions of their own leadership. Nonetheless, teachers are certainly one important source for evaluating leadership practices and improvement processes in the school.

Moreover, we found that the relationship of leadership and ending learning achievement was fully rather than partially mediated. That is, we discerned no direct effect of leadership on learning in any of our empirical tests. As noted in a long series of scholarly reviews (e.g. Bossert *et al.*, 1982; Bridges, 1982; Creemers and Kyriakides, 2008; Erickson, 1967; Hallinger, 2011; Hallinger and Heck, 1996, 1998; Leithwood *et al.*, 2004, 2006; Robinson *et al.*, 2008; Scheerens, 2012; Witziers *et al.*, 2003), tests of direct-effects models have yielded few results that are statistically significant, theoretical compelling, or of practical utility. The results counter continuing assertions that school leadership has a direct, measurable effect on student learning (e.g. Nettles and Herrington, 2007; Silva *et al.*, 2011). We further assert that the approach to modeling the linkages between leadership, teaching, and learning employed in this study constitutes one of the strongest explicit tests of this hypothesis conducted to date (see also Heck and Hallinger, 2009, 2010).

Third, our results offer a new insight by illuminating the role that leadership can play in shaping the instructional environment and coordinating the instructional practices of teachers. More specifically, in this study leadership enhanced the direct effect of teachers on student math achievement. Notably, our school-level constructs (i.e. leadership and instructional environment) were developed from detailed information obtained from teachers about their schools' leadership and instructional processes. These school-level factors predicted levels of individual teacher effectiveness on student learning within schools. Moreover, when we aggregated teacher effectiveness scores within schools, we found superior improvement in student performance in schools where teaching, on average, was more effective and where there was less variation across teacher in their levels of effectiveness.

Thus, we conclude that consistency in teaching quality also matters. Or stated differently, student opportunity to learn with a succession of effective teachers in schools with a more consistent instructional environment and stronger instructionally focussed leadership, appears to produce "compounding benefits." This calls attention to the role of leadership not only in increasing "average teaching effectiveness" but also in reducing variability in effectiveness across teachers who comprise the faculty within a given school.

This finding highlights the role of instructional leaders in enhancing consistency in teacher effectiveness. It implies the potential utility of various leadership strategies including engagement in recruitment and selection of teachers, organization of coaching and teacher professional development, and modeling "cultural values and norms" that emphasize and support quality teaching and learning (Hallinger, 2011; Hallinger and Heck, 1998, 2010; Heck and Hallinger, 2009, 2010; Leithwood *et al.*, 2004, 2006, 2010a, b; Marks and Printy, 2003; May and Supovitz, 2011; Robinson, 2011; Robinson *et al.*, 2008; Printy *et al.*, 2009; Sebastian and Allensworth, 2012).

Moreover, linking this conclusion to the finding that leadership for learning is a "fully mediated" relationship (Preacher, 2011) reinforces the recommendation that school leaders should focus on "intermediate targets" that have proven impact on student learning (Creemers and Kyriakides, 2008; Hallinger and Heck, 1998; Leithwood *et al.*, 2004, 2006). The "paths" investigated in this paper (i.e. instructional environment and teacher effectiveness), as well as paths discussed elsewhere (e.g. Creemers and Kyriakides, 2008; Day *et al.*, 2011; Hallinger, 2011; Hallinger and Heck, 1998; Hattie, 2009; Heck and Hallinger, 2009, 2010; Leithwood *et al.*, 2004, 2006, 2010a, b; Mulford and Silins, 2003; Printy *et al.*, 2009; Robinson *et al.*, 2008; Sebastian and Allensworth, 2012) represent reasonable targets. Nonetheless, we emphasize that the conditions which describe the context of a particular school also determine which paths or targets offer the greatest leverage in actual school improvement efforts.

Almost 20 years ago, we highlighted the need to advance our understanding of the "paths" through which leadership contributes to school improvement (Hallinger and Heck, 1996, 1998). During the ensuing period, scholars have sought to address this goal through literature review (e.g. Hallinger, 2003, 2011; Leithwood *et al.*, 2004, 2006; Robinson *et al.*, 2008; Scheerens, 2012; Witziers *et al.*, 2003), cross-sectional studies of school leadership effects on learning (e.g. Day *et al.*, 2011; Geijsel *et al.*, 1999, 2003; Hallinger *et al.*, 1996; Ko *et al.*, 2012; Kruger *et al.*, 2007; Leithwood and Jantzi, 2000, 2010a, b; Marks and Printy, 2003; Sebastian and Allensworth, 2012; Walker and Ko, 2011; Wiley, 2001), and longitudinal studies of leadership and school improvement (e.g. Hallinger and Heck, 2010, 2011; Heck and Hallinger, 2009, 2010; Mulford and Silins, 2009; Thoonen *et al.*, 2012).

This study sought to demonstrate the theoretical and practical advantages to be gained from studying cross-level interactions among school- and classroom-level factors proposed to contribute to school improvement. We believe that the findings advance our understanding of a key "classroom path" that links leadership and learning. Moreover, we also suggest that the multilevel, cross-classified modeling of longitudinal data employed in this research offers scholars another useful methodology for examining school improvement processes and outcomes.

Notes

1. In a multilevel mediated-effects relationship, we would propose that school leadership actions (X) affect student classroom achievement (Y) indirectly by creating conditions that

impact teacher's classroom effectiveness as the mediating effect. Teacher effectiveness, in turn, affects student learning. We tested for this effect separately in preliminary analyses but found that teachers' perceptions of school leadership were not statistically significantly related to individual value-added teacher estimates.

2. We note that in previous studies we have examined variability in student, teacher, and parent responses to school leadership and classroom practices and found that our school level models work very similarly across different school role groups. The levels of perceptions systematically vary, with teachers' generally most positive and parents least positive about school processes (e.g. Heck and Hallinger, 2009).

3. For the baseline model, we assume the observed status of outcome Y at time t for individual i, cross-classified in class j_1 and j_2 (which are represented in parentheses to indicate the cross classification) in school k is captured in an intercept parameter (π_0), a growth parameter (π_1), plus random error (ε) in estimating the outcome:

$$Y_{it(j1j2)k} = \pi_{0i(j1j2)k} + \pi_{1i(j1j2)k}a_{it} + \varepsilon_{it(j1j2)k}. \tag{1}$$

In this equation a_{it} represents the growth slope that specifies the change in individual i at time t between the repeated math measurements. We consider this level fixed due to the added complexity of estimating the models for random student-level effects.

At the teacher level, the cross-classified classroom structure is modeled:

$$\pi_{0i(j1j2)k} = \beta_{00k} + u_{0j1k} + u_{0j2k}, \tag{2}$$

,

$$\pi_{1i(j1j2)k} = \beta_{10k}. \tag{3}$$

At this level, in the intercept model (Equation 2) there is a random term added for each classroom. A third random effect could be added in Equation (2) representing the effect of being in a particular years 1 and 2 classroom (i.e. $U_{0j1\mathrm{x}j2}$). However, it is generally set to zero, since it can be difficult to separate the student-level variance from the variance of the classroom interaction (Raudenbush and Bryk, 2002). We tested whether student growth varied across the second classroom but found that it did not. We therefore specified the student growth rate (π_1) as fixed at the classroom level in Equation 3.

Between schools, there are two random parameters indicating that students' ending status intercepts and growth slopes vary across schools:

$$\beta_{00k} = \gamma_{000} + \upsilon_{0k} \tag{4}$$

,

$$\beta_{10k} = \gamma_{100} + \upsilon_{1k}. \tag{5}$$

A set of school predictors was added at their appropriate levels of the model (as shown in Figure 1). Finally, at the last step, the mediated and cross-level relationships can be investigated. We explain these investigations in further details in the results section.

4. The standard error of the indirect effect can be estimated from the two direct effects a and b (between X and M and between M and Y) and their corresponding t-values as follows:

$$se = \frac{\hat{a}\hat{b}\sqrt{t_{\hat{a}}^2 + t_{\hat{b}}^2}}{t_{\hat{a}}t_{\hat{b}}}$$

5. We note that the moderating effects of school's instructional environment on the effectiveness slopes were similarly statistically significant for the second teacher and positive, but not statistically significant, for the first teacher.
6. It should be noted the in the earlier study, we only tested for change in "school-level relationships" rather than in cross-level (i.e. school to classroom) relationships.
7. Here "stability" refers to the use of a common set of instruments and data collection process over a sufficient period of time. For example, the dataset that we used in the earlier study was no longer available due to changes in the pattern and focus of data collection employed by the particular state education department. The technical requirements for conducting the type of research demonstrated in the current study are considerable. For example, the data set must link students with specific teachers as they move through their educational careers in order to model these cross-level changes.

References

Baron, R.M. and Kenny, D.A. (1986), "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations", *Journal of Personality and Social Psychology*, Vol. 51 No. 6, pp. 1173-1182.

Bidwell, C.E. (1965), "The school as a formal organization", in Marsh, J.G. (Ed.), *Handbook of Organizations*, Rand McNally, Chicago, IL, pp. 972-1022.

Bidwell, C.E. (2001), "Analyzing schools as organizations: long-term permanence and short-term change", *Sociology of Education*, Vol. 74 No. 5, pp. 100-114.

Bloom, H.S., Hill, C.J., Black, A.R. and Lipsey, M.L. (2008), "Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions", MDRC working papers on Research Methodology, MDRC, October, New York, NY.

Borman, G. and Kimball, S. (2005), "Teacher quality and educational equality: do teachers with higher standards-based evaluation ratings close student achievement gaps?", *The Elementary School journal*, Vol. 106 No. 1, pp. 3-20.

Bossert, S., Dwyer, D., Rowan, B. and Lee, G. (1982), "The instructional management role of the principal", *Educational Administration Quarterly*, Vol. 18 No. 3, pp. 34-64.

Bridges, E. (1982), "Research on the school administrator: the state-of-the-art, 1967-1980", *Educational Administration Quarterly*, Vol. 18 No. 3, pp. 12-33.

Coelli, M. and Green, D.A. (2012), "Leadership effects: school principals and student outcomes", *Economics of Education Review*, Vol. 31 No. 1, pp. 92-109.

Creemers, B. and Kyriakides, L. (2008), *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*, Routledge, New York, NY.

Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., Brown, E., Ahtaridou, E. and Kingston, A. (2011), *Successful School Leadership: Linking Learning and Achievement*, Open University Press, Berkshire.

De Fraine, Van Damme, J. and Onghena, P. (2002), "Accountability of schools and teachers: what should be taken into account?", *European Educational Research Journal*, Vol. 1 No. 3, pp. 403-427.

Erickson, D. (1967), "The school administrator", *Review of Educational Research*, Vol. 37 No. 4, pp. 417-432.

Geijsel, F., Sleegers, P. and Van den Berg, R. (1999), "Transformational leadership and the implementation of large-scale innovation programs", *Journal of Educational Administration*, Vol. 37 No. 4, pp. 309-328.

Geijsel, F., Sleegers, P., Leithwood, K. and Jantzi, D. (2003), "Transformational leadership effects on teachers' commitment and effort toward school reform", *Journal of Educational Administration*, Vol. 41 No. 3, pp. 228-256.

Goldhaber, D.D. (2002), "The mystery of good teaching", *Education Next*, Vol. 2 No. 1, pp. 50-55.

Good, T.L. and Brophy, J.E. (1986), "Teacher behavior and student achievement", in Whittrock, M.C. (Ed.), *Handbook of Research on Teaching*, 3rd ed., McMillan, New York, NY, pp. 328-375.

Hallinger, P. (2003), "Leading educational change: reflections on the practice of instructional and transformational leadership", *Cambridge Journal of Education*, Vol. 33 No. 3, pp. 329-351.

Hallinger, P. (2011), "Leadership for learning: lessons from 40 years of empirical research", *Journal of Educational Administration*, Vol. 49 No. 2, pp. 125-142.

Hallinger, P. and Heck, R.H. (1996), "Reassessing the principal's role in school effectiveness: a review of empirical research, 1980-1995", *Educational Administration Quarterly*, Vol. 32 No. 1, pp. 5-44.

Hallinger, P. and Heck, R.H. (1998), "Exploring the principal's contribution to school effectiveness: 1980-1995", *School Effectiveness and School Improvement*, Vol. 9 No. 2, pp. 157-191.

Hallinger, P. and Heck, R. (2011), "Conceptual and methodological issues in studying school leadership effects as a reciprocal process", *School Effectiveness and School Improvement*, Vol. 22 No. 2, pp. 149-173.

Hallinger, P. and Heck, R.H. (2010), "Collaborative leadership and school improvement: understanding the impact on school capacity and student learning", *School Leadership & Management*, Vol. 30 No. 2, pp. 95-110.

Hallinger, P., Bickman, L. and Davis, K. (1996), "School context, principal leadership and student achievement", *The Elementary School Journal*, Vol. 96 No. 5, pp. 498-518.

Hallinger, P., Heck, R. and Murphy, J. (2014), "Teacher evaluation and school improvement: an analysis of the evidence", *Educational Assessment Evaluation and Accountability*, Vol. 26 No. 1, pp. 5-28.

Hanushek, E. (1971), "Teacher characteristics and gains in student achievement: estimation using micro data", *American Economic Review*, Vol. 61 No. 2, pp. 280-288.

Hattie, J.A.C. (2009), *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Routledge, London.

Hawley, W. and Rosenholtz, S. (1984), "Good schools: what research says about improving school achievement", *Peabody Journal of Education*, Vol. 61 No. 4, pp. 117-124.

Heck, R.H. (2009), "Teacher effectiveness and student achievement: investigating a multilevel cross-classified model", *Journal of Educational Administration*, Vol. 47 No. 2, pp. 227-249.

Heck, R.H. and Hallinger, P. (2009), "Assessing the contribution of distributed leadership to school improvement and growth in math achievement", *American Educational Research Journal*, Vol. 46 No. 3, pp. 626-658.

Heck, R.H. and Hallinger, P. (2010), "Collaborative leadership effects on school improvement: integrating unidirectional- and reciprocal-effects models", *The Elementary School Journal*, Vol. 111 No. 2, pp. 226-252.

Heck, R.H., Larson, T. and Marcoulides, G.A. (1990), "Principal instructional leadership and school achievement: validation of a causal model", *Educational Administration Quarterly*, Vol. 26 No. 2, pp. 94-125.

Kleine-Kracht, P. (1993), "Indirect instructional leadership: an administrator's choice", *Educational Administration Quarterly*, Vol. 18 No. 4, pp. 1-29.

Kline, R.B. (2004), *Principles and Practices of Structural Equation Modeling*, 2nd ed., Guilford Press, New York, NY.

Ko, J., Hallinger, P. and Walker, A. (2012), "Exploring school improvement in Hong Kong secondary schools", *Peabody Journal of Education*, Vol. 87 No. 2, pp. 216-234.

Kruger, M., Witziers, B. and Sleegers, P. (2007), "The impact of school leadership on school level factors: validation of a causal model", *School Effectiveness and School Improvement*, Vol. 18 No. 1, pp. 1-20.

Kupermintz, H. (2003), "Teacher effects and teacher effectiveness: a validity investigation of the Tennessee value added assessment system", *Educational Evaluation and Policy Analysis*, Vol. 25 No. 3, pp. 287-298.

Kyriakides, L., Creemers, B., Antoniou, P. and Demetriou, D. (2009), "A synthesis of studies searching for school factors: implications for theory and research", *British Research Journal*, Vol. 36 No. 1, pp. 1-24.

Lazarsfeld, P. and Fiske, M. (1938), "The 'panel' as a new tool for measuring opinion", *Public Opinion Quarterly*, Vol. 2 No. 4, pp. 596-612.

Lee, V.E. and Bryk, A.S. (1989), "A multilevel model of the social distribution of high school achievement", *Sociology of Education*, Vol. 62 No. 3, pp. 172-192.

Leithwood, K. (2001), "School leadership in the context of accountability policies", *International Journal of Leadership in Education*, Vol. 4 No. 3, pp. 217-235.

Leithwood, K. and Jantzi, D. (2000), "The effects of transformational leadership on organizational conditions and student engagement with the school", *Journal of Educational Administration*, Vol. 38 No. 2, pp. 112-129.

Leithwood, K., Patten, S. and Jantzi, D. (2010b), "Testing a conception of how school leadership influences student learning", *Educational Administration Quarterly*, Vol. 46 No. 5, pp. 671-706.

Leithwood, K., Anderson, S., Mascall, B. and Strauss, T. (2010a), "School leaders' influences on student learning: the four paths", in Bush, T., Bell, L. and Middlewood, D. (Eds), *The Principles of Educational Leadership and Management*, 2nd ed., Sage, London, pp. 13-30.

Leithwood, K., Louis, K., Anderson, S. and Wahlsttom, K. (2004), "Review of research: how leadership influences student learning", available at: www.wallacefoundation.org/NR/rdonlyres/E3BCCFA5-A88B-45D3-8E27-B973732283C9/0/ReviewofResearchLearningFromLeadership.pdf (accessed December 19, 2007).

Leithwood, K., Day, C., Sammons, P., Harris, A. and Hopkins, D. (2006), *Seven Strong Claims about Successful School Leadership*, National College of School Leadership, Nottingham.

Luyten, H., Visscher, A. and Witziers, B. (2005), "School effectiveness research: from a review of the criticism to recommendations for further development", *School Effectiveness and School Improvement*, Vol. 16 No. 3, pp. 249-279.

McCaffrey, D.F., Lockwood, J.R., Koretz, D., Louis, T.A. and Hamilton, L. (2004), "Models for value-added modeling of teacher effects", *Journal of Educational and Behavioral Statistics*, Vol. 29 No. 1, pp. 67-101.

MacKinnon, D.P. (2008), *Introduction to Statistical Mediation Analysis*, Taylor and Francis, New York, NY.

Marks, H. and Printy, S. (2003), "Principal leadership and school performance: an integration of transformational and instructional leadership", *Educational Administration Quarterly*, Vol. 39 No. 3, pp. 370-397.

Marsh, H.W. and 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*, Vol. 1 No. 2, pp. 133-163.

May, H. and Supovitz, J.A. (2011), "The scope of principal efforts to improve instruction", *Educational Administration Quarterly*, Vol. 47 No. 2, pp. 332-352.

Mendro, R., Jordan, H., Gomez, E., Anderson, M. and Bembry, K. (1998), "Longitudinal teacher effects on student achievement and their relation to school and project evaluation", paper presented at the annual meeting of the American Educational Research Association, San Diego, CA, April.

Milanowski, A. (2004), "The relationship between teacher performance evaluation scores and student achievement: evidence from Cincinnati", *Peabody Journal of Education*, Vol. 79 No. 4, pp. 33-53.

Mulford, B. and Silins, H. (2003), "Leadership for organisational learning and improved student outcomes – what do we know?", *Cambridge Journal of Education*, Vol. 33 No. 2, pp. 175-195.

Mulford, B. and Silins, H. (2009), "Revised models and conceptualization of successful school principalship in Tasmania", in Mulford, B. and Edmunds, B. (Eds), *Successful School Principalship in Tasmania*, Faculty of Education, University of Tasmania, Launceston, pp. 157-183.

Murnane, R. (1975), *The Impact of School Resources on the Learning of Inner City Children*, Ballinger, Cambridge, MA.

Murphy, J., Hallinger, P. and Heck, R.H. (2013), "Teacher evaluation: a case of the missing clothes?", *Educational Researcher*, Vol. 42 No. 6, pp. 349-354.

Muth, B.O. (1994), "Multilevel covariance structure analysis", *Sociological methods & Research*, Vol. 22 No. 3, pp. 376-398.

Nettles, S. and Herrington, C. (2007), "Revisiting the importance of the direct effects of school leadership on student achievement: the implications for school improvement policy", *Peabody Journal of Education*, Vol. 82 No. 4, pp. 724-736.

Opdenakker, M.C. and Van Damme, J. (2007), "Do school context, student composition, and school leadership affect school practice and outcomes in secondary education?", *British Educational Research Journal*, Vol. 33 No. 2, pp. 179-206.

Oud, J.H.L. (2002), "Continuous time modeling of the cross-lagged panel design", *Kwantitatieve Methoden*, Vol. 23 No. 69, pp. 1-26.

Pitner, N. (1988), "The study of administrator effects and effectiveness", in Boyan, N. (Ed.), *Handbook of Research in Educational Administration*, Longman, New York, NY, pp. 106-132.

Printy, S.M., Marks, H.M. and Bowers, A.J. (2009), "Integrated leadership: how principals and teachers share transformational and instructional influence", *Journal of School Leadership*, Vol. 19 No. 5, pp. 504-532.

Preacher, K.J. (2011), "Multilevel SEM strategies for evaluating mediation in three-level data", *Multivariate Behavioral Research*, Vol. 46 No. 4, pp. 691-731.

Raudenbush, S.W. and Bryk, A.S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage, Thousand Oaks, CA.

Reynolds, D., Teddlie, C., Hopkins, D. and Stringfield, S. (2000), "Linking school effectiveness and school improvement", in Teddlie, C. and Reynolds, D. (Eds), *The International Handbook of School Effectiveness Research*, Falmer Press, London, pp. 206-231.

Robinson, V.M. (2011), *Student-Centered Leadership*, Jossey-Bass, San Francisco, CA.

Robinson, V.M., Lloyd, C. and Rowe, K. (2008), "The impact of leadership on student outcomes: an analysis of the differential effects of leadership types", *Educational Administration Quarterly*, Vol. 44 No. 5, pp. 564-588.

Rosenshine, B. (1983), "Teaching functions in instructional programs", *Elementary School Journal*, Vol. 83 No. 4, pp. 335-351.

Rowan, B., Correnti, R. and Miller, R.J. (2002), "What large-scale survey research tells us about teacher effects on student achievement: insights from the *Prospects* study of primary schools", *Teachers College Record*, Vol. 104 No. 8, pp. 1525-1567.

Sanders, W. and Horn, S.P. (1998), "Research findings from the Tennessee value-added assessment system (TVASS) database: implications for educational evaluation and research", *Journal of Personnel Evaluation in Education*, Vol. 12 No. 3, pp. 247-256.

Scheerens, J. (2012), *School Leadership Effects Revisited*, Springer, Dordrecht.

Sebastian, J. and Allensworth, E. (2012), "The influence of principal leadership on classroom instruction and student learning: a study of mediated pathways to learning", *Educational Administration Quarterly*, Vol. 48 No. 4, pp. 626-663.

Seidel, T. and Shavelson, R.J. (2007), "Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results", *Review of Educational Research*, Vol. 77 No. 4, pp. 454-499.

Seltzer, M., Choi, K. and Thum, Y.M. (2003), "Examining relationships between where students start and how rapidly they progress: using new developments in growth modeling to gain insight into the distribution of achievement within schools", *Educational Evaluation and Policy Analysis*, Vol. 25 No. 3, pp. 263-286.

Silva, J., White, G. and Yoshida, R. (2011), "The direct effects of principal–student discussions on eighth grade students' gains in reading achievement: an experimental study", *Educational Administration Quarterly*, Vol. 47 No. 5, pp. 772-793.

Southworth, G. (2002), "Instructional leadership in schools: reflections and empirical evidence", *School Leadership and Management*, Vol. 22 No. 1, pp. 73-92.

Stoll, L. and Fink, D. (1996), *Changing our Schools*, Open University Press, London.

Teddlie, C. and Stringfield, S. (1993), *Schools Make a Difference: Lessons Learned From a 10-Year Study of School Effects*, Teachers College Press, New York, NY.

Thoonen, E., Sleegers, P., Oorta, F. and Peetsmaa, T. (2012), "Building school-wide capacity for improvement: the role of leadership, school organizational conditions, and teacher factors", *School Effectiveness and School Improvement*, Vol. 23 No. 4, pp. 441-460.

Urick, A. and Bowers, A.J. (2011), "What influences principals' perceptions of academic climate? A nationally representative study of the direct effects of perception on climate", *Leadership and Policy in Schools*, Vol. 10 No. 3, pp. 322-348.

Walker, A. and Ko, J. (2011), "Principal leadership in an era of accountability: a perspective from the Hong Kong context", *School Leadership and Management*, Vol. 31 No. 4, pp. 369-392.

Wiley, S. (2001), "Contextual effects on student achievement: school leadership and professional community", *Journal of Educational Change*, Vol. 2 No. 1, pp. 1-33.

Willms, J. (1992), *Monitoring School Performance: A Guide for Educators*, Falmer Press, London.

Witziers, B., Bosker, R. and Kruger, M. (2003), "Educational leadership and student achievement: the elusive search for an association", *Educational Administration Quarterly*, Vol. 39 No. 3, pp. 398-425.

Further reading

Hill, P. and Rowe, K. (1996), "Multilevel modeling in school effectiveness research", *School Effectiveness and School Improvement*, Vol. 7 No. 1, pp. 1-34.

Rothstein, J. (2009), "Teacher quality in educational production: tracking, decay and student achievement", Working Paper No. 14442, National Bureau of Economic Research, Washington, DC.

About the authors

Dr Ronald H. Heck holds the Dai Ho Chun Endowed Chair in Education at the U. of Hawaii-Manoa. He received his PhD in Education from the University of California at Santa Barbara.

His research focusses on the relationship between school leadership and school improvement. He has also written on research methodology. Dr Ronald H. Heck is the corresponding author and can be contacted at: rheck@hawaii.edu

Dr Philip Hallinger is the Chair Professor and Director of the Asia Pacific Centre for Leadership and Change at the Hong Kong Institute of Education. His research focusses on school leadership effects, leadership development and problem-based learning. He received his EdD in Administration and Policy Analysis from the Stanford University.