Paper submitted September 2010 to
School Effectiveness and School Improvement

Exploring the Journey of School Improvement: Classifying and Analyzing Patterns of Change in School Improvement Processes and Learning Outcomes

Philip Hallinger

Hong Kong Institute of Education

Hallinger@gmail.com

And

Ronald Heck

University of Hawaii - Manoa

Abstract

The policy goal of enhancing the academic achievement of students in under-performing schools has taken on greater urgency in recent years. While this has increased attention on leading improvement in 'turn-around' schools, this article suggests that it is equally important to meet the challenge of providing guidance for evidence-based improvement in schools more generally. This study examines a longitudinal data set that describes the performance of 193 elementary schools in the USA over a four-year period of time. This research sought first to determine if it was possible to classify schools in terms of different patterns of growth in their learning outcomes over time. Then the research sought to link these patterns of growth, or school improvement, to both features of the school context and changes in school leadership and academic capacity. The study found that schools could be successfully classified according to several predominant patterns of growth in learning, and that these could be linked to features of the school context as well as to changes in specific alterable school-level conditions. The study's longitudinal design focusing on growth and change enabled the research to extend prior school effectiveness findings into the domain of school improvement.

> As Ronald Edmonds often said, we know far more about the features that characterize an effective school than we know about how a school became effective in the first place. Why, then, do we try to force schools that we don't like, to resemble schools that we do like, by employing means that have little to do with the evolution of the kind of schools that we like? (Barth, 1986, p. 294)

Twenty-five years hence, Barth's question continues to echo in the halls of academia as scholars seek to understand, interpret, and convey the applicability of their research to the improvement of practice in schools. For example, even as researchers point to substantial progress in understanding the means by which leadership contributes to learning in schools (e.g., author id. ref.; Leithwood, Patten & Jantzi, 2010; Marks & Printy, 2003; Robinson, Lloyd & Rowe, 2008), we remain highly constrained in our ability to answer a practical question recently posed to us by a school principal: 'Given what you know about leadership for learning, where would you advise me to put my effort as a school leader in order to gain the greatest improvement in learning for students at my school" (H.S. Shui, personal communication, March 29, 2010)?

While this is a legitimate question, the answer requires a contextualization of research findings that goes beyond the limits of the current literature in school improvement. For example, early studies in this domain focused on leadership in poor urban elementary schools that had been judged to be instructionally effective (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1977; Edmonds, 1979; Purkey & Smith, 1983). Yet, as suggested by Barth (1986) and others, these schools were 'atypical' in a variety of important respects that could be traced to the context. Therefore, the findings,

while inspiring and important, had limitations as a guide for practice in schools that operated in different circumstances (Cuban, 1984).

In subsequent years, researchers expanded their studies of school leadership effects to a much broader range of schools (author id. ref.; Bell, Bolam, & Cubillo, 2003; Southworth, 2002; Witziers, Bosker, & Kruger, 2003). However, only recently has empirical research begun to hone in on explicit linkages between patterns of successful leadership practice and the *context of schools* (e,g., author id. ref.; Day, Sammons, Leithwood, Hopkins, Harris, Gu, & Brown, 2010; Leithwood, Harris, & Strauss, 2010; Murphy & Meyers, 2008; Opdenakker & Van Damme, 2007). By *context* we refer to environmental and organizational conditions that moderate the school's capacity for improving student learning. This paper takes as a central premise that school leaders must take these contextual conditions into account as they work towards improvements in school performance.

This report builds upon a set of prior studies in which we explored the relationship between leadership and school improvement (author id. ref.). These studies sought to understand how changes in leadership were related to changes in academic capacity and growth in student learning over a four-year period of time in a sample of 200 elementary schools. These analyses confirmed the validity of mediated- and reciprocal-effects perspectives on school leadership in which changes in leadership and academic capacity together positively impacted growth in student achievement (author id. ref.).

In the current study, we drill down into the same dataset to explore patterns of improvement among these elementary schools. We use as a starting point the 'change

trajectories' (e.g., stable, improving, declining) of individual schools (author id. ref.). We then investigate whether it is possible to identify 'latent classes', or common patterns among the improvement trajectories of the schools. Drawing upon Jackson's (2000) metaphor of school improvement as a journey, we were interested in:

1. Identifying whether schools can be classified in terms of different patterns of school-level growth in student learning over time (i.e., on different journeys),
2. Determining whether these journeys are moderated by selected features of the school context identified as important from empirical and practical perspectives,
3. Linking patterns of growth in student achievement with change in alterable school-level conditions (i.e., leadership, academic improvement capacity).

The significance of this study lies in establishing whether it is possible to identify patterns in the improvement trajectories that describe the actual performance of schools over a substantial period of time. The ability to classify schools in this manner would represent progress towards the development of evidence-based practice in school improvement. This knowledge base would seek to link patterns of leadership and school improvement practice to growth in learning for schools located at different points in their improvement journeys. This would begin to address both Roland Barth's (1986) challenge and the practical question posed above by our principal.

Theoretical Perspective

In this section of the paper we begin by reviewing theoretical perspectives and methodological approaches used by researchers in the study of school improvement. This builds the rationale for the approach taken in this particular study. Then we present the

conceptual model that guides our this study and review relevant findings from other empirical research.

Conceptualizing the Study School Improvement

Researchers interested in the study of school improvement have employed a variety of theoretical perspectives. Scholars have referred to the knowledge base on effective teaching and learning (Creemers, 1994; Creemers & Kyriakides, 2008; Hattie, 2009; Mortimore, 1993), teacher development (Fullan & Hargreaves, 1992; Joyce & Showers, 2002), effective schools (Brookover et al., 1977; Edmonds, 1979; Purkey & Smith, 1983), and school leadership (author id. ref.; Bossert et al., 1982; Day et al., 2010; Leithwood et al., 2004; Opdenakker & Van Damme, 2007; Robinson et al., 2008; Southworth, 2002) on the assumption that these literatures would provide insights into the nature of classroom and school-level practices that represent high leverage foci for school improvement.

Other scholars have framed the study of school improvement in terms of processes associated with change on the assumption that school improvement is a form of organizational change and development. These efforts have focused on personal (Bridges & Bridges, 2009; Evans, 1996; Maurer, 1996; Rogers, 2003; Schön, 1983), organizational (Drucker, 1996; Kanter, 1995; Kotter, 1996; O'Toole, 1995; Weick, 1976), and educational change (Cuban, 1990; Firestone & Corbett, 1988; Fullan, 2006, 2007; Hall & Hord, 2002; Sleegers, Geijsel & Van den Berg, 2002). This line of inquiry has sought to describe and analyze processes that could impact successful change in schools regardless of the particular focus (e.g., teaching method, curriculum, program, whole school improvement). Variants on this approach have also focused more narrowly

on studying change from the perspectives of curriculum, innovation, and program implementation in schools (Berman & McLaughlin, l978; Crandall, Eiseman & Louis, 1986; Darling Hammond, 2006; Fullan & Pomfret, 1977; Hall & Hord, 2002; McLaughlin, 1990; van den Berg & Ros, 1999).

Another popular lens for exploring school improvement emerged from the literatures on school and organizational culture (Deal & Peterson, 2009; Saphier & King, 1985; Sarason, 1982; Schein, 1996). Later efforts that evolved from this perspective have been linked to the construct of learning organizations, which highlights the systemic nature of change in schools (Leithwood & Louis, 2000; Mulford & Silins, 2003, 2009; Senge, 1990; Silins & Mulford, in press). Finally, a more narrowly focused literature has also evolved around the study of school improvement as a domain in and of itself (e.g., Edmonds, 1982; Foster, 2005; Harris, 2006; Hawley & Rosenholtz, 1984; Jackson, 2000; Nicolaidou, & Ainscow, 2005; Purkey & Smith, 1985; Reynolds, Teddlie, Hopkins, & Stringfield, 2000; Stoll & Fink, 1996). Scholars have, however, had difficulty reconciling these literatures because they are often grounded in different assumptions and research traditions (Ouston, 1999). Thus, extant attempts to synthesize these diverse theoretical perspectives on school improvement (e.g., Fullan, 2003, 2006, 2007; Leithwood, Patten & Jantzi, 2010; Mulford & Silins, 2003; Stoll & Fink, 1996) have yet to result in a single overarching theoretical perspective.

Despite this limitation, we do note substantial progress both in analyzing and elaborating on important principles and underlying structures and processes associated with successful change and improvement in schools (e.g., Fullan, 2007). For example, research clearly highlights the importance of leadership during the change process, both

as a catalyst and agent of support (e.g., author id, 2003; Hall & Hord, 2002). Moreover, in recent years scholars have begun to differentiate more clearly the nature of leadership that may be needed during the turnaround stage in schools facing special measures (Duke, 2004; Leithwood, Harris & Strauss, 2010; Murphy & Meyers, 2008; Nettles & Herrington, 2007).

Other research has begun to examine the relationship between the school's context, leadership and improvement strategies of schools during change. Recent empirical studies conducted by Opdenakker and Van Damme (2007), Day and colleagues (2010), and Author (id. ref.) have sought to examine the relationship between contextual conditions of schools and school leadership. For example, Day and colleagues (2010) identified different patterns of successful school leadership that corresponded to four sequential, developmental stages in the journey of school improvement. They termed these stages: 1) coming out of special measures, 2) taking ownership, 3) developing creativity, and 4) everyone a leader. In this study, the improvement of school performance was explicitly linked to changing patterns of leadership and the development of school capacity for improvement (Day et al., 2010).

These findings reprise a long tradition in the organizational behavior literature that has sought to link different approaches to leadership to features of the organizational context (e.g., Fiedler, 1967; Hersey & Blanchard, 1977). Although earlier conceptualizations emphasized the need for more direction and control in organizational contexts that demonstrated low capacity and performance (e.g., Hersey & Blanchard, 1977), the more recent literature on turn-around schools suggests that capacity development and team development begin from day one (Duke, 2004; Leithwood, Harris

& Strauss, 2010; Murphy & Meyers, 2008). Then, as suggested by Day and colleagues (2010), leadership becomes increasingly collaborative as school capacity for improvement becomes more broadly distributed. Thus, we have begun to see a more nuanced understanding of the meaning of 'strong leadership' in the more recent school improvement literature.

These theoretical perspectives guided the framing and selection of variables in the current study. For example, the study incorporated contextual conditions as a set of moderating variables and employed a longitudinal design so as to capture change an unfolding process. Finally, the study centered on leadership and school capacity as central constructs proposed to impact growth in student learning over time.

Methodological Perspectives on the Study of School Improvement

At the outset of this report, we reprised Barth's (1986) critique of school improvement policy research because we believe that the state-of-the-art in the literature on school improvement continues to rely far too heavily on the interpretation of findings that do not explicitly describe change and improvement in schools. Case studies have been useful at describing strategies and challenges in trying to 'turn around' schools (e.g., Nicholaidou & Ainscow, 2005) and generating descriptions of what occurs during efforts to improve schools in other specific contexts (e.g., author id. ref; Harris, 2006; Jackson, 2000; Stoll & Fink, 1996). Nonetheless, case study findings cannot be generalized, and a research and development strategy predicated on building a knowledge base from case studies alone will be both laborious and of limited validity.

School effectiveness surveys have been useful at broadening our understanding of factors associated with effective schools and schools that have demonstrated improvements (Brookover et al., 1977; Mortimore, 1993; Reynolds et al., 2000). However, improvement, by definition, entails change in the state of schools over time. *Cross-sectional surveys* only offer a one-time snapshot of the state of the school's performance and, therefore, can make only limited contributions to understanding core issues underlying the processes associated with school improvement (author id. ref.; Luyten et al., 2005; Mulford & Silins, 2003).

A third approach to developing the knowledge base in school improvement has entailed conceptual analysis and review of related literatures. However, this approach to mapping knowledge accumulation, while useful, relies upon findings from the first two empirical approaches. It therefore, cannot advance the knowledge base beyond the confines of existing knowledge.

In sum, we assert that a robust strategy for developing substantive knowledge concerning valid strategies and practices for *school improvement* must include the analysis of longitudinal data that describe the performance of relatively large numbers of schools over time (author id. ref.; Day et al., 2010; Mulford & Silins, 2009). In the absence of longitudinal data, the field is unlikely to progress in developing a more sophisticated understanding of the nature of school improvement processes and outcomes in schools operating under different conditions (e.g., different cultures, community types, school levels). Opdenakker and Van Damme (2007) framed this argument with respect to their own study of school improvement.

> While in the past researchers often concentrated on one specific potentially effectiveness-enhancing factor (e.g. school leadership), or investigated the influence of several factors on outcomes without taking into account the possibility that factors relate to each other, nowadays there are calls to pay attention to the interrelatedness of factors, the direct and indirect effects of factors, the mediated effects of factors, and to use *time-ordered modeling* [emphasis added] procedures like path analysis or structural equation modeling. So, a plea for more complex models is made (see, for example, Witziers et al., 2003). (pp. 179-180)

From a practical perspective, longitudinal data are needed if we are to gain a better understanding of the patterns of change that occur across schools during the 'journey of school improvement' (Jackson, 2000). The paucity of longitudinal data with which to conduct research on improvement across large numbers of schools has, however, until recently stalled progress in identifying and understanding patterns of change in the improvement of schools. We note that the necessity of using longitudinal data also extends to studies of leadership impact on school improvement. Here Ogawa and Bossert (1995) state the case as follows.

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

Indeed, as noted above, it is only relatively recently that researchers have begun to conduct longitudinal studies of school improvement on a scale that is capable of linking key improvement foci to conditions of the school's context and 'stage of improvement' (e.g., author id. ref.; Day et al., 2010; Mulford & Silins, 2003, 2009). This work is being facilitated by a trend whereby governmental bodies in various parts of the world have begun to develop substantial longitudinal datasets on school processes and outcomes. We suggest that to the extent that policymakers wish to gain valid insights into school improvement processes, they should support this trend since the collection of this type of data is generally beyond the means of individual or even teams of researchers.

With these comments in mind, we highlight the fact that this study incorporated data collected annually from 193 schools over a four-year period of time. While longer time durations are, of course, preferable, four years should be sufficient to identify whether medium-term patterns of school growth can be detected. Moreover, we suggest that four years represents a meaningful duration of time since it matches the average tenure of school principals in many parts of the world.

A Proposed Model of Leadership and School Improvement

Figure 1 portrays the general model that has guided our research. It builds on current theoretical models which propose that successful schools engage in intentional strategies and actions to improve learning environments and teaching practices (e.g., Creemers, 1994; Edmonds, 1982; Hattie, 2009; Kyriakides & Creemers, 2010; Leithwood et al., in press; Mortimore, 1993; Opdenakker & Van Damme, 2007; Robinson et al., 2008). The model further suggests that influences on student learning accrue at multiple levels of the school organization. In this case, we focus on school context and process

relationships that are assumed to influence the environment in which learning and teaching take place.

The main variables in this study included leadership, academic (or instructional) improvement capacity, and student achievement. The relationships among these variables are assumed to change in predictable ways over time. Finally, these changes are proposed to link to certain contextual features of schools. For this last proposition, we did not specify the nature of contextual impact, but rather framed the study so that it could be accounted for and described.

The model proposes that factors at the school level have both direct and indirect effects on student achievement, not only because they influence student achievement at the school level, but also because they directly and indirectly influence the composition of classrooms as well as teaching and learning in classrooms. We define leadership as a collaborative learning-directed construct that is distributed among various individuals in different school roles, focused on fostering conditions that support effective teaching and learning and build capacity for professional learning and change (author id. ref.; Gronn, 2002; Kyriakides & Creemers, 2008; Robinson et al., 2008). It emphasizes *governance* that empowers others and encourages broad participation and responsibility, building collaboration in school improvement *decisions*, and broad participation in *evaluating* the school's academic development. School leadership is proposed to achieve its effects on academic outcomes indirectly through building the school's professional capacity and by maintaining a focus on improvements in teaching and learning. For this study, we define this as building the school's academic (or instructional) capacity (e.g., Creemers & Kyriakides, 2008; Opdenakker & Van Damme, 2007). This model assumes that changes

in the underlying learning-directed leadership and academic capacity constructs at the school level manifest themselves in unmeasured changes in teachers' classroom practices and students' experiences, which, in turn, are responsible for student growth in math, along with student background variables.

In this study, we represent this multilevel relationship in the *within-schools* portion of the model in Figure 1 with a dotted oval and arrows, since we did not include direct classroom measures. We acknowledge that school-level aggregates employed can ignore wide variations in the conditions of learning and teaching that may be very important at the classroom level (Creemers & Kyriakides, 2008). We note, however, that in previous three-level (i.e., school, classrooms, students) tests of this proposed theoretical model, we found collaborative learning-directed leadership focused on building academic capacity moderated subsequent teacher effectiveness at the classroom level, which, in turn, influenced student growth in reading and math (author id ref., 2010).

[Insert Figure 1 about here]

We illustrate the proposed impact of these school-level, latent relationships on classrooms and students with a broad arrow extending from the school level to the classroom level and (by association) to the individual student level of the data hierarchy. Even though we do not measure *classroom changes* directly in this analysis, we assume that changes to teacher classroom behaviors will be responsible for changes observed in student growth rates (Creemers, 1994; Hill & Rowe, 1996; Lee & Bryk, 1989; Mortimore, 1993; Robinson et al., 2008). Because we utilized information from teachers and triangulated it with similar information from students and parents, we believe the

data provide a reasonable means to test the proposed conceptual model (see for example, author ref, id.).

In this paper, as noted, we focus not so much on direct efforts to improve classroom teaching behavior, but rather on school-level efforts to improve the learning environment. This includes not only student learning, but also teacher professional development, teacher collaboration, student support systems, resource allocation, and academic focus and expectations (author id. ref.; Kyriakides & Creemers, 2010; Leithwood et al., 2004, in press; Robinson et al., 2008). The school's focus on teaching and learning practices and the relative presence of these factors we term the school's capacity for academic improvement, or academic capacity.

As suggested in Figure 1, this research focused on the relationships observed over time between several sets of variables and growth in student learning. These included features of the school's context, collaborative learning-directed leadership, and academic capacity. The model seeks to portray the changing relationship among these moderating and mediating variables over time as they related to leadership and learning. It should be noted that this conceptual model incorporates both static and dynamic feature of the relevant variables in a *change model* of school improvement. This general model has been tested in prior analyses (author id. ref.) whose findings can be broadly summarized as follows:

- Change in collaborative learning-directed leadership positively impacted growth in student learning by shaping the school's academic capacity.
- Collaborative learning-directed leadership and capacity building were mutually reinforcing in their effects on each other, and exercised a

cumulative impact on student learning in both reading and math at the elementary school level.

- It was further noted that academic capacity produced a larger effect on leadership than leadership upon capacity at each measurement occasion over a four year period. This led to the conclusion that although learning-directed leadership may be a necessary catalyst for school improvement, its impact on learning is still less substantial than that of the school's academic capacity.

- Schools in which the same principal was present over the four-year period of the study demonstrated stronger growth in learning-directed leadership and stronger academic capacity at the end of the four year period.

- The initial status of schools' academic capacity was not related to subsequent growth, which implies that schools could improve regardless of where they were located in their journey of school improvement.

The purpose of this paper is to extend these previous findings. More specifically, we seek to explore in greater detail patterns in the trajectories that schools move along in their attempts to improve alterable conditions at different stages in their journeys of school improvement.

Method

This study employed a non-experimental, post-hoc, longitudinal design (Campbell & Stanley, 1966). Although superior to cross-sectional designs for this type of research, longitudinal studies cannot fully resolve the direction of causality between variables, (Cook, 2002). Major threats to validity in longitudinal, non-experimental research include possible omitted variables, inappropriate measurement of the constructs over time, possible selection bias effects, and subject dropouts. Omitted variables (e.g., school-level covariates) and inappropriate measurement of latent constructs by their observed indicators are common sources of misspecification that can produce misleading results in

structural equation models (Bentler & Bonett, 1980). For example, an individual's reported involvement in school decision making may, or may not, adequately capture a key aspect of leadership; and even if it does, the way the individual's reply is coded into a score may bias its exact meaning (Bentler & Bonett, 1980). We took a number of preliminary steps to ensure that our constructs were measured properly over time.[i]

Although changes in leadership and school capacity building may be related to increased student growth rates, they do not provide complete protection against a selection-bias argument. For example, teachers may perceive improvement capacity more positively in schools that achieve at high levels over long periods of time. Additional waves of data would enable a more robust time-series examination of the causal linkages in the proposed change model over a longer period of time. We discuss preliminary analyses to address possible effects of subject drop out (missing data) in more detail in the following section.

Data Source

From the database of all public elementary schools in a western US state, we drew a random sample of 193 elementary schools. A longitudinal cohort of 13,391third-grade students within these schools (Mean = 70.81, SD = 43.37) was followed over a three-year period. Background data were as follows: female, 49%; participation in federal free/reduced lunch program, 45%; receiving English language services, 7%; receiving special education services, 11%; minority, 50%, and changed schools, 16%. Achievement data from the student cohort were collected in years 2, 3 and 4 (i.e., corresponding to their third through fifth grade years).

Information about school processes was collected school surveys about school processes. The survey is administered by the educational system at regular cycles in each school to all certified staff, grade five students, and a random sample of parents (i.e., about 20% across grade levels in each school). Because teachers are well positioned to understand the school's curriculum, instructional expectations and routines, and are in contact with students and parents regularly, we decided to capture potential changes in leadership and academic processes using the surveys given to each school's teachers on three occasions (year 1, year 3, year 4) over the four-year period of the study. Return rates for the three data collection occasions were 73.4% (N = 3,911), 78.6% (N = 4,152), and 76.2% (N = 4,055) respectively. To examine possible differences in our model due to perceptions of other school stakeholders, we also re-ran our analyses with the parent and student data (with similar results) to extend our model's generalizability.

Where surveys are repeated over time with a high level of consistency between items, the measures may be used to estimate changes in a population, even if the samples may be different on each occasion (Davies, 1994). This design is referred to as a longitudinal panel study, as opposed to a cohort study, because the survey cannot be linked to the same set of teachers over time due to various reasons (e.g., lack of teacher identifiers, teacher turnover). One concern in longitudinal studies is that there is generally some dropping out over time. Rubin (1976) introduced the notion of the distribution of 'missingness' as a way to classify the conditions under which missing data can be ignored.

Little and Rubin (2002) distinguish between data missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Missing

completely at random implies that the data do not depend on other variables in the model (e.g., covariates or outcomes). For example, this is generally the case when a random sample is taken from a population. MAR implies that the missing data may depend on other variables in the model. In both cases, however, the missing case is assumed to be independent of the value on the unobserved (missing) variable (Hox, 2010). MNAR suggests that probability of dropping out is related to responses at the time of dropping out. In longitudinal panel studies, however, we typically have information about individuals who participate (e.g., mobility, perceptions about school processes, student outcomes) from previous occasions. In such cases, it is generally reasonable to assume MAR, conditional on those variables, which also includes scores on the outcomes at earlier times (Hox, 2010; Schafer, 2005).

We conducted follow-up analyses to estimate the extent to which missing data on teachers and students might affect our results. Regarding teachers, first, we estimated that teacher turnover at schools in the sample averaged about 8% per year during the years of our study. The average number of teachers per school was 20.5 teachers. On average, therefore, 2 teachers left their school each year. We next investigated how school conditions influenced patterns of teacher mobility over time. We noted that the total set of school conditions (i.e., student composition, student achievement, enrollment size) and staffing conditions (i.e., teacher, principal stability, teacher experience) in our model contributed little in explaining school variability in teacher turnover over the years of the study (i.e., about 1%, not tabled).[ii] Schafter (1997) concludes that assuming MAR is reasonable in such circumstances.

Third, regarding mobility in our longitudinal student cohort, 14 percent of the students entered the school system after the first year of the study (not tabled), and 16 percent changed schools. Preliminarily, we developed dummy variables regarding whether students entered the school system late and whether they changed schools. We then examined whether school-level and student level variables were associated with students' likelihood to enter the system late or to change schools. We found there were similar patterns of missing data observed for both conditions we examined.[iii] We therefore concluded there were no systematic biases related to differential patterns of missing data by students who were stable in their schools over time versus students who moved or entered the cohort late.

One advantage of the LCA approach is that missing data and student mobility can be incorporated directly into the analysis, which reduces parameter bias that could result from eliminating these cases (Peugh & Enders, 2004). When using maximum likelihood (ML) estimation, there is no need to delete incomplete subject data. Estimation is based on available data points, and subjects do not need to have complete data. As Peugh and Enders (2004) note, partial data actually contribute to the estimation of the model's parameters by implying probable values for missing scores via the correlations among variables. Expectation maximization, a common method for obtaining ML estimates with incomplete data, treats the model parameters (rather than the data points themselves) as missing values to be estimated and borrows information from the existing data at successive iterations until differences between covariance matrices generated are trivial.

Variables in the Model

In our following analyses we make use of several different sets of explanatory variables from our proposed model. We next define them and describe them in further detail. Our conceptual model of school change processes incorporates focuses on school characteristics concerned with the context of the school, student and staff composition, and school practice.

Context, student composition, and staffing variables. For the purposes of this study, we selected several context, student composition, and staffing variables that have been identified as factors that could moderate the practice and impact of school leadership. Features of small schools appear to favor enhanced growth in student learning (Opdenakker & Van Damme, 2007; Southworth, 2002). *Enrollment size* (i.e., defined as the number of students enrolled for the full year) is likely a proxy measure of more complex processes that provide benefits to students in smaller school settings (e.g., enhanced social relationships including teacher-student interactions; school organization and practices).

Previous research on educational effectiveness identified the inequitable distribution of student learning resulting from student background (e.g., language background, gender, socioeconomic status) within and between schools (e.g., Creemers & Kyriakides, 2008; Hill & Rowe, 1996; Lee & Bryk, 1989). Students' social backgrounds influence grouping strategies as well as access to curriculum and quality teaching (Darling-Hammond, 2000). Student composition has been found to have broad effects on academic expectations, curriculum organization, grouping, and teacher behavior (Lee & Bryk, 1989; Lee & Burkham, 2003; Opdenakker & Van Damme, 2007).

We defined *student composition* as a weighted factor of the percentages of students receiving free/reduced lunch (a measure of average student SES), English language services, and minority by race/ethnicity.

Teacher certification (i.e., percentage of students fully meeting state licensing requirements) information and *staff stability* (i.e., percentage of teachers at the school over the previous five-year period; whether the same principal was present over the length of the study) are also important to include as control variables because previous research has found that schools in some locations have a difficult time hiring and retaining quality faculty and administrators (Darling Hammond, 2006). Research further confirmed that interactions between the school's environment and its internal organization form a context in which school leadership is exercised (e.g., author id. ref.; Bossert et al., 1982; Leithwood et al., 2004; Ogawa & Bossert, 1995; Teddlie, Stringfield, & Reynolds, 2000).

Collaborative learning-directed leadership. We defined collaborative learning-directed leadership as a latent construct that is proposed to drive development of the school's academic capacity. Four assumptions frame the study's definition of leadership. First, the practice of leadership involves developing a vision for change and then motivating and enabling people to achieve the vision (author id. ref.; Leithwood et al., 2004; Marks & Printy, 2003). Second, leadership in schools tends to be distributed and collaborative; therefore, its measurement should not be limited to the actions of those in formal management roles (Day et al., 2006; Gronn, 2002; Leithwood et al., 2004). Third, effective school leadership creates conditions that support effective teaching and learning (author id. ref.; Fullan, 2006; Leithwood, Patten, & Jantzi, 2010; Marks & Printy, 2003;

Robinson et al., 2008; Wiley, 2001). Fourth, leadership that increases the school's capacity for academic improvement will impact student achievement positively (Bell et al., 2003; Fullan, 2006; Lee & Bryk, 1989; Leithwood et al., 2004, in press; Mulford & Silins, 2003; Stoll & Fink, 1996).

Given its centrality to this study, the second assumption concerning collaborative leadership requires additional elaboration. Although researchers have traditionally emphasized leadership exercised by those holding hierarchical positions, scholars have become increasingly interested in conceptions that highlight the distribution of leadership among individuals holding a wider range of organizational roles (Day et al., 2006; Gronn, 2002). This collaborative perspective has received particular emphasis in the literature on school leadership due to characteristics of schools as 'team-based' (Day et al., 2006; Gronn, 2002), 'loosely-coupled' organizations (Ogawa & Bossert, 1995; Weick, 1976). In combination with recent changes originating in the institutional environment (e.g., increasing accountability, work intensification, role differentiation), this has led to a growing *prescriptive*, as well as *descriptive* interest in, the sharing or distribution of leadership in schools (Gronn, 2002; Leithwood et al., 2004).

Scholars suggest that the impact of collaborative leadership in schools is achieved through improved communication of mission and goals, better alignment of resources and structures to support students, more active engaged professional learning among staff, and the ability to maintain a focus on innovations in teaching and learning by those responsible for implementation (Fullan, 2006; Gronn, 2002; Leithwood et al., 2004; Robinson et al., 2008).

For this study, collaborative school leadership was conceived as a form of

'learning-directed leadership' (author ref, 2003). It was measured by a subscale (α = 0.82) describing teacher perceptions of leadership exercised from a variety of sources within the school (e.g., principal, grade-level heads, teachers, and community representatives on the school's leadership council), which is focused on sustained school improvement, support for shared governance, involvement in resource allocation, and evaluation of school-based improvement efforts. The survey items were designed to reflect three specific aspects of school leadership: *school improvement* (i.e., To what extent does school leadership: Make decisions to facilitate actions that focus the energies of the school on student achievement and school-wide learner outcomes; Empower staff and students; Encourage commitment, participation and shared accountability for student learning?); *school governance* (Adopt governance guidelines which are consistent with the school's purpose and support the achievement of the state standards and the school-wide learner outcomes?); and *resource management and development* (Allocate available resources in a manner that sustains the school program and are used to carry out the school's purpose; Use assessment results the basis for the allocation and use of resources?). Factor scores describing the measurement of the leadership factor on each of the three occasions (summarized in the results section) were saved and used to define the LCA model of school leadership.

We emphasize that this scale assessed the 'strength' of learning-directed leadership rather than the 'degree of collaboration.' Higher scores on this scale do not indicate stronger collaboration on leadership within the school. Rather they should be interpreted as an indication that the school's leadership was more consistently and coherently directed towards supporting teaching and learning to improve academic

outcomes.

School improvement in academic capacity. A substantial body of research has found that leadership effects in schools are mediated by the school's academic and social organization (author id. ref.; Leithwood et al., 2004). For the purposes of this study, we refer to this mediating factor as the school's capacity for academic improvement. This factor is defined from a set of discrete variables that have emerged from several decades of research on school effectiveness and improvement (Teddlie & Reynolds, 2000).

The specific observed indicators that comprised this latent factor included the quality of 1) the school's implementation of the state's curricular standards, 2) sustained focus on school academic improvement, 3) support for students, 4) professional capacity of the school, 5) school communication, 6) stakeholder involvement, and 7) student safety and well being. Tapping into this underlying process with multiple measures and time points provides a valuable way to monitor evolving organizational work structures proximal to student learning across a large number of cases.

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

- *Standards emphasis and implementation* (α = 0.91). School's educational programs are aligned to the State content and performance standards; teaching

and learning activities are focused on helping students meet the State content and performance standards; school prepares students well for the next school; students and parents are informed about what students are expected to learn; school has high academic and performance standards for students; classroom instruction includes active participation of students; curriculum and instructional strategies emphasize higher-level thinking and problem solving; instructional time is flexible and organized to support learning; teachers provide a variety of ways for students to show what they have learned; students learn to assess their own progress and set their own learning goals; students are provided with multiple ways to show how well they have learned; homework assignments are appropriate, productive, and reflective of adopted learning standards; assessment results are used to plan and adjust instruction;

- *Focused and sustained action on improvement* (α = 0.83). School clearly communicates goals to staff, parents and students; vision and purpose are translated into appropriate educational programs for children; school seeks ways to improve its programs and activities that promote student achievement; teachers know what the school learner outcomes are; teachers expect high quality work; school's vision is regularly reviewed with involvement of all stakeholder groups; changes in curriculum materials and instructional practices are coordinated school-wide and I am involved in the school improvement process;
- *Quality of student support* (α = 0.85). Standards exist for student behavior; discipline problems are handled quickly and fairly; school environment

supports learning; open communication exists among administrators, teachers, staff, and parents; teachers feel safe at school; teachers and staff care about students; administrators, teachers, and staff treat each other with respect; I provide students with extra help when they need it; programs meet special needs of students; school reviews support services offered to students;

- *Professional capacity of the school* (α = 0.80). Teachers are well qualified for assignments and responsibilities; leadership and staff are committed to school's purpose; staff development is systematic, coordinated, and focused on standards-based education; systematic evaluation is in place;
- *School communication* (α = 0.88). School employs a wide range of strategies to ensure parent involvement; open communication among staff; open communication exists between school staff and parents; school responds to parent concerns; school keeps parents informed; I encourage and welcome parents to come to my classroom);
- *Stakeholder involvement* (α = 0.80). Parents participate in important decisions about their children's education; school involves parents in classrooms such as tutoring students or checking homework; school encourages parent involvement in a variety of ways; and
- *Student safety and well-being* (α = 0.82). The school is orderly and supports learning, school staff shows that they respect and care about students, students can receive extra help and support when needed.

Reliability and validity. Various forms of the survey instrument employed in this study have been used previously research studies. This research has shown the subscales

defining the school improvement capacity factor to be reliable and valid. The internal consistency estimates obtained meet common standards for survey research (i.e., with α coefficients of 0.80 or above). Validity of the scales has been assessed for both face and predictive validity. For example, prior studies found a significant relationship between the quality of schools' improvement capacity measured by the survey and sixth-grade student achievement in reading and math (author id. ref.), Standardized effects for explaining student achievement levels ranged from 0.08 to 0.34 across five different years of achievement data and from 0.22 to 0.31 on student growth rates across multiple student cohorts. Prior studies therefore provide consistent evidence of the instrument's reliability and validity.

Math achievement. The math test used in the study was constructed to measure state-developed math content standards. The test consisted of constructed-response items and standardized test items from the Stanford Achievement Test (Edition 9). The test assesses student learning in five strands (number and operation; measurement; geometry and spatial sense; patterns, functions and algebra; and data analysis, statistics, and probability) consisting of 52 items. Student scores (re-scaled to range from 100 to 500) considered patterns of right, wrong, and omitted responses over successive years and were equated across the three years to enable the measurement of academic growth.

Descriptive Statistics

Table 1 provides information about the school-level and student background variables used on our models. The average school size was 469 students (SD = 213). The percentages of students receiving English Language services (ELL), low SES, and minority were combined through principal components analysis to create a school

indicator of student composition (M = -0.03, SD = 0.95). Regarding staffing, about one-third of the principals (31%) were in the same school over the four plus years of the study. Nearly 60% of teachers were stable over the previous five-year period in which the study took place. Eighty-four percent of the teachers were fully qualified according to state teacher licensing standards. Regarding initial leadership and school improvement capacity, Year 1 scores reflect the equating of the scores to 0 in order to establish measurement invariance over time. Successive Year 3 and Year 4 scores in Table 1 indicate positive growth, although the metric of the subsequent factor scores has no inherent meaning (Raykov & Marcoulides, 2006).

[Insert Table 1 about Here]

Data Analysis

Our analyses proceeded in two primary phases. First, we used multilevel *latent change analysis* (LCA), which facilitates the examination of changes in leadership and organizational processes within the same analysis as changes in student learning outcomes. In LCA, repeated observations on individuals over time (y_t) can be expressed as a measurement model where the intercept and growth latent factors are measured by the multiple indicators of *y*. The intercept factors representing the constructs are defined to represent initial levels of each factor, which is accomplished by setting each factor loading to 1.0. The growth factors were defined to incorporate possible nonlinearity in the growth trajectories. This can be accomplished by fixing the first measurement occasion factor loading to 0, the second occasion to 1 (representing linear change), and letting the third factor loading be estimated by the software. The size of the estimated

factor loading then determines whether the growth trajectory has a nonlinear shape (Raykov & Marcoulides, 2006).

Our tests of the proposed model presented in Figure 1 exploited this capability of LCA to examine the relationships among multiple variables at multiple levels over time within a single simultaneous model. We defined latent factors measuring students' true initial levels of academic achievement in math and their *true* change (or growth) over the length of the study. Students' observed academic outcomes were measured using math achievement scaled scores collected on three occasions (end of grades 3, 4, and 5) in Years 2, 3, and 4. We examined changes in leadership and school processes by measuring these variables through teacher surveys given in Year 1, Year 3, and Year 4. For each measurement occasion, teacher surveys were administered from one year (Year 1) to several months (Years 3 and 4) before students' math progress was assessed.

As suggested in Figure 1, we also included a range of student background (e.g., gender, SES, ethnicity, language background) and school context controls (e.g., school size, student composition, staff stability). The inclusion of these variables helps in interpreting the final relationships between our three main constructs of interest (i.e., changes in leadership, improvement capacity, and growth in student achievement). We determined that model fit the data well and confirmed that *changes* in leadership focused on schools' capacity for improving their instruction and student systems were positively related to *growth* in student learning over the four year period of our study (author id. ref.).

Second, after examining changes in our sample schools over the four-year period, we attempted to isolate emergent groups of schools that experienced similar patterns of

student growth in math and then identify school-level variables that might explain differences in the patterns of school growth we observed. This type of investigation is conducted by defining a categorical latent variable that represents underlying *mixtures* (or latent classes) in a population, where population membership is not known ahead of time but, rather, is inferred from the data (Muthén, 2001). Mixture models refer to models with a categorical latent variable that represents mixtures of subpopulations. The purpose of this type of modeling is to estimate the number and size of latent classes in the mixture and then assign membership to the latent classes. This part of our analysis focused on investigating possible heterogeneity in the math trajectories found in the population of schools in order to explore whether these might be related to key moderating school variables (e.g., school size, social composition, principal stability, teacher staffing), as well as school leadership and academic capacity. Our basic latent class model (which also includes the school portion of the latent change model) is summarized in Figure 2.

[Insert Figure 2 about here]

The model suggests that the latent categorical variable representing an unknown number of latent classes (or subpopulations with similar growth over time) can capture similarities between schools in terms of their growth trajectories. It is assumed these latent classes represented by the categorical latent variable have different means over time, which can result in different improvement trajectories for groups of schools. The model also relates the latent classes to leadership and capacity conditions, as well as other context covariates that might be responsible for observed differences in the subgroup means. After identifying subgroups of schools, we provide several follow-up analyses

where we continued to examine variables that might explain differences in school patterns of growth in math.

Results

Our first proposition suggested that there could be identifiable patterns among schools in the growth trajectories associated with school-level math achievement of the student cohorts over the four-year period. Tests of our proposed model in Figure 1 suggested that three latent classes provided the best model fit to the data. We represent the differences in estimated means at three time intervals (i.e., years). The average trajectories over time of these three classes of schools are shown in Figure 3.

As the figure suggests, the three classes of schools start out at very different points in terms of their estimated initial math scores. The 'typical' set of schools is indicated by Class 1 (n = 161). Class 1 schools started out close to the sample mean and grew about 12 scaled-score points per year over Years 2 through 4. The second and third classes are represented by a smaller number of schools (n = 16 in each class) with distinctly different growth trajectories. Class 2 schools (High Growth) started with very low estimated initial school-level math scores but made relatively high growth over time. Indeed, Class 2 schools made impressive gains that succeeded in narrowing the 'achievement gap' when compared with the initial achievement of Class 1 and Class 3 schools. Class 3 schools started out somewhere between Class 1 and 3 schools and achieved moderate growth over time. Thus, we conclude that it is possible to develop a classification of schools according to their patterns of growth in learning over time.

[Insert Figure 3 about here]

After identifying classes of growth trajectories, our second proposition sought to determine if the three classes of schools were also characterized by different patterns of change in relevant context and school-level process variables. Three main findings stand out in terms of our latent mixture model analysis (i.e., from Figure 2).

- When compared with Class 2 schools, Class 1 and Class 3 schools were larger in enrollment and comprised of significantly less challenging student compositions.
- When compared with Class 2 schools, Class 1 and Class 3 schools demonstrated significantly higher "average" academic capacity over the course of the four years.
- When compared with Class 2 schools, Class 1 and Class 3 schools demonstrated significantly lower levels of learning-directed leadership over time.

Next we employed discriminant analysis to determine whether the more complete set of predictors in the framework (see Figure 1) would be useful in classifying the individual schools into 'correct' latent classes. With three categories for the outcome, two possible functions could classify schools according to their contexts and processes. These functions were significant when analyzed together; moreover, when the first function was removed, the second remained significant ($\chi^2 = 8.06$, $p < .05$).

[Insert Table 2 about Here]

Next we wish to highlight the variables that were most responsible for classifying schools from our larger model in Table 2. The standardized function coefficients indicate the relative strength of each variable in explaining the classification results for each function. The first function (along the horizontal axis in the figure) separates Class 1 schools (Typical) from Class 2 (High Growth) and Class 3 (Moderate Growth) schools. For this function, student composition (.89), school academic capacity (.77), perceptions

of weaker leadership (-.44), and higher percentages of teachers that met state certification requirements (0.31) contributed most substantially towards classifying schools.

The second function (along the vertical axis in the figure), distinguishes Class 2 (High Growth) from Class 3 (Moderate Growth) schools. As Table 2 suggests, teachers in Class 2 schools perceived stronger learning-directed leadership than teachers in Class 3 schools (0.67). Although teachers in Class 2 schools perceived lower academic capacity (-.66), Class 2 schools had higher percentages of teachers who met state certification requirements and greater staff stability (0.22) than teachers in Class 3 schools. Principal stability was also notable in distinguishing Class 2 and Class 3 schools. Moreover, having the same principal (versus having different principals) over the four-year period seems to have contributed more positively to growth in math in Class 2 schools, as compared with Class 3 or Class 1 schools. In sum, it was possible to classify 85% of the 193 schools correctly according to membership in one of these latent classes, and the results suggest that differences in schools' academic capacity and leadership practices contribute meaningfully to understanding variations in schools' achievement trajectories.

[Insert Figure 4 about here]

Figure 4 further illustrates the interplay of *changes* in collaborative learning-directed leadership and academic capacity over the four year period for the three classes of school growth in math. We sought additional evidence of differences in the relationship between changes in these two primary constructs. The curve represents the "best fitting" line (i.e., in this case a curvilinear, or quadratic, shape) describing this relationship within each class of schools. In Class 1 schools, the relationship between learning-directed leadership and academic capacity was almost linear over time. As

collaborative learning-directed leadership strengthens so does academic capacity, though we note that, consistent with the finding reported earlier, in these schools academic capacity tended to be stronger than leadership as the schools progressed over time. As Figure 4, indicates the relationship accounted for 9% of the variance between them.

In contrast, the relationship between changes in school leadership and academic capacity appear to be more consequential among Class 2 schools as compared with Class 1 or Class 3 schools. Note the positive relationship between change in leadership and capacity building is stronger in this set of schools (as suggested by the larger R^2 coefficient of 0.27). We noted earlier, Class 2 schools were initially lower in terms of capacity and, like Class 3 schools, also had challenging student circumstances. Keeping in mind that Class 3 schools achieved more moderate growth in math over time than Class 2 schools, in Class 3 schools, relationship between change in school leadership and capacity building was considerably weaker than in Class 2 schools ($R^2 = 0.17$), but also stronger than in Class 1 schools. Follow-up multivariate analyses suggested the relationship between change in leadership and change in capacity building was significantly different across latent classes ($F_{4,187} = 4.46$, $p < .05$), even after controlling for student composition, enrollment, principal stability, and teacher variables (not tabled).

We also noted that in schools with the same principal over time teacher-perceived changes in leadership and academic capacity were significantly larger (0.19, 0.21, respectively, $p < .05$) than in schools where there was principal turnover. Finally, we also found that principal stability interacted with the latent classes ($F_{2,187} = 19.07$, $p < .01$) in defining changes in academic capacity over time. More specifically, assuming that no change in capacity over time would yield a mean change = 0, Class 1 schools that had

experienced principal turnover had a mean change in academic capacity of -0.16 SD (i.e., teachers perceived they actually declined in capacity over time). In contrast, perceived changes in academic capacity were considerably stronger in Class 2 schools (Mean = 0.42 SD) and Class 3 schools (Mean = 0.10 SD) where the same principal was present over the four years of the study. We do remind readers, however, that they represent descriptions in relatively small groups of schools that demonstrated considerably different growth patterns from those of the majority of schools in the sample. It is clear, however, that stronger collaborative learning-directed leadership and increasing academic capacity (along with increased principal stability) characterized these schools making greater than average growth in math achievement over time.

Conclusions

Across the globe, governments have undertaken strategies to restructure schools that fail to meet minimum standards of performance in student achievement. This is readily apparent, for example, in the USA and the UK where sustained levels of poor performance can trigger extreme measures including the replacement of principals and teachers, or even closure of schools. In this new policy climate, schools are now routinely required to formulate school improvement or development plans that focus on student learning outcomes. Yet, shaping an evidence-based plan that meet the needs of a particular school's context remains a challenge (author id. ref.; Day et al., 2010; Jackson, 2000).

To date, researchers have only been able to offer assistance at a fairly general level of abstraction. For example, a decade ago, Hargreaves and Fullan (1998) suggested the limits of what research could offer:

> There is no ready answer to the "how" question. Singular recipes oversimplify what it will take to bring about change in your own situation. Even when you know what research and published advice tell you, no one can prescribe exactly how to apply to your particular school and all the unique problems, opportunities and peculiarities it contains. (p. 106)

As we noted earlier in this report, the impact of research on school improvement has been limited by the field's reliance on relatively weak research designs. Knowledge gained from cross-sectional surveys and case studies has largely failed to cohere into a sound knowledge base that links improvement practices with contexts and outcomes. As observed by Opdenakker & Van Damme (2007):

> Quite often relationships between school characteristics and school outcomes are studied without taking into account possible relations and influences between school characteristics. Effects of school context, school leadership, student composition and school practice on school outcomes are studied without paying attention to possible relations between school characteristics and mediator effects. The consequences of this are inconsistencies in research results between studies and a rather limited understanding of the educational effectiveness of schools. (p. 194)

In our view, these 'conditions' include not only important features of the context (e.g., student composition, school size, school level), but also the school's location (i.e., current status) and trajectory (i.e., stable, declining, improving) on its 'journey' of school improvement. These factors combine to create a 'unique set of improvement challenges' for each school. However, we wish to suggest that properly designed research can provide results that shed light on how leaders respond to these conditions in order to

develop the capacity of their schools to improve (e.g., see Day et al., 2010; Leithwood , Patten & Jantzi, 2010; Mulford & Silins, 2009; Silins & Mulford, in press; Opdenakker & Van Damme, 2007).

With this in mind, we suggest that this study makes an incremental contribution to the literature on school improvement by demonstrating that it is possible to classify schools according to where they are located in their improvement journeys. This means that we need not treat every school's context as completely unique. Furthermore, the study has taken first steps towards linking these classifications to patterns of practice. This is another dimension of what Hallinger and Heck (1996) and Leithwood and colleagues (Leithwood, Patten & Jantzi, 2010) terms identifying the "paths" by which leadership impacts learning. Of course, the current study has only set down markers for these paths. We trust that future research will expand on this initial effort through studies that link context to leadership, other school conditions and improvement in school performance over time.

These results reinforce the long-standing empirical finding that leadership acts as a catalyst for school improvement, both by initiating change and shaping a coherent focus on learning in schools (author id. ref., 2003; Bossert et al., 1982; Edmonds, 1982; Firestone, & Corbett, 1988; Fullan, 2006; Leithwood & Montgomery, 1982). The findings suggest that the development of academic capacity results from a collective effort, even from the early stages in the school improvement journey (se also Duke, 2004). More specifically, our research indicates that the relationship between leadership and academic capacity is best described as one of mutual influence or reciprocity (see

also author ref, in press). Reciprocity emerged as we sought to model the relationship between school leadership and academic capacity in the three sets of schools.

Strong learning-directed, collaborative leadership appeared to be an important factor, or catalyst, for change in the capacity of the low performing schools to improve. We surmise that in school contexts where academic capacity is weak, leadership from the principal and others is essential to bringing about improvement. The data also suggested that both principal stability and principal support for collaborative leadership may have been particularly important in sustaining change in the high growth schools. The nature of the measurement tools in this study did not, however, allow us to decompose the leadership contributions that were due to the principal and others.

In contrast, the other two groups of schools (i.e., Classes 1 and 3) began with stronger performance outcomes and higher academic capacity, but weaker scores on the leadership measure. We hypothesize that as academic capacity develops in a school it may become a substitute for leadership as the sustaining engine for improvement (Kerr & Jermier, 1978). This suggests that, over time, leadership may diffuse through the organization, transforming from an individual characteristic (e.g., the principal), to an attribute of a team, and finally into an organizational property (Ogawa & Bossert 1995). This mirrors the fourth stage of school improvement identified by Day and colleagues (2010), which they termed “Everyone a Leader.”

We wish to note once again that similar findings have emerged out of other recent medium to large-scale longitudinal studies of leadership and school improvement (author ref; Day et al., 2010; Mulford & Silins, 2009; Opdenakker & Van Damme, 2007; Silins & Mulford, 2010). This evidence of progress reinforces our earlier assertions concerning

the utility of longitudinal designs in a long-term program of research on school improvement. Indeed, we suggest that these studies typify a new generation of longitudinal research on school improvement.

Our own results provide an initial demonstration of the potential utility of longitudinal modeling of school change in performance over time. We suggest further exploration of this approach as a means of narrowing down the range of 'school improvement journeys' that schools undertake. This line of inquiry could, in the future, result in more fine-grained advice for policy and practice concerning the formulation of contextualized school improvement strategies. We assert that the next generation of school improvement studies should maintain this focus since, at the end of the day, contributions to policy and practice must be able to provide not only 'lists of relevant factors', but also their relative importance for different types of schools located at different stages in their school improvement journeys.

References

Barth, R. (1986). On sheep and goats and school reform. *Phi Delta Kappan, 68*(4), 293-296.

Bell, L., Bolam, R., & Cubillo, L. (2003). *A systematic review of the impact of school head teachers and principals on student outcomes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

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

Berman, P. & McLaughlin, M. (l978). Implementation of educational innovations. *Educational Forum, 40*, 47-370.

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

Bridges, W., & Bridges, S. (2009). *Managing transitions: Making the most of change*. Philadelphia: Da Capo Press.

Brookover, W., Beady, C., Flood, P., Schweitzer, J. & Wisenbaker, J. (1977). *Schools can make a difference*. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED145034)

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

Crandall, D., Eiseman, J., & Louis, K.S. (1986). Strategic planning issues that bear on the success of school improvement efforts. *Educational Administration Quarterly, 22*(3), 21-53.

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

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1), 3-13.

Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review, 54*(2), 128-151.

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

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

Day, C., Sammons, P., Leithwood, K., Hopkins, D., Harris, A., Gu, Q., & Brown, E. (2010). *Ten strong claims about successful school leadership*. Nottingham, UK: The National College for School Leadership.

Deal, T., & Peterson, K.D. (2009). Shaping school culture: Pitfalls, paradoxes, and promises. San Francisco: Jossey Bass.

Drucker, P. (1995). *Managing in a time of great change*. New York: Talley House, Dutton.

Duke, D. (2004). The turnaround principal: High-stakes leadership, *Principal, 84*(1), 12-23.

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

Edmonds, R. (1982). Programs of school improvement: An overview. *Educational Leadership, 12*(4), 125-128.

Evans, R. (1996). *The human side of school change*. San Francisco: Jossey Bass.

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

Firestone, W. A. & Corbett, H. D. (1988). Planned educational change. In N. Boyan (Ed.) *Handbook of research on educational administration* (pp. 321-340). New York: Longman.

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

Fullan, M. (2003). *Change forces with a vengeance*. London: Routledge-Falmer.

Fullan, M. (2007). *The new meaning of educational change* (4th Edition). New York: Teachers College Press.

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

Fullan, M., & Hargreaves, A. (1992). *Teacher development and educational change*. London/Washington: Falmer

Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research, 47*(1), 335-397.

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

Hall, G. & Hord, S. (2002). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn & Bacon.

Hargreaves, A., & Fullan, M. (1998). *What's worth fighting for out there.* New York: Teachers College Press.

Harris, A. (2006). Leading change in schools in difficulty. *Journal of Educational Change, 7*(1/2), 9-18.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

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

Hersey, P., & Blanchard, K. H. (1977). *Management of organizational behavior: Utilizing human resources* (3rd ed.). Paramus, New Jersey, Prentice Hall.

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

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

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: ASCD.

Kanter, R.M. (1995). *World class: Thriving locally in the global economy*. New York: Simon and Schuster.

Kerr, S, & Jermier, J. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational behavior and human performance. 22*(3), 375-403.

Kotter, J. (1996). *Leading change*. Boston, HBS Press.

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

Kyriakides, L., & Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183–205.

Lee, V., & Bryk, A. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*(3), 172-192.

Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal, 40*(3), 353-394.

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

Leithwood, K., Harris, A., & Strauss, T. (2010). *Leading school turnaround.* San Francisco: Jossey Bass.

Leithwood, K., & Louis, K.S. (2000). *Understanding schools as intelligent systems*. New York: JAI Press.

Leithwood, K., Louis, K. S., Anderson, S. & Wahlsttom, K. (2004). Review of research: How leadership influences student learning. Wallace Foundation. Downloaded from http://www.wallacefoundation.org/NR/rdonlyres/E3BCCFA5-A88B-45D3-

8E27-B973732283C9/0/ReviewofResearchLearningFromLeadership.pdf on December 19, 2007.

Leithwood, K., & Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research, 52*(3), 309-339.

Leithwood, K., Patten, S., & Jantzi, D. (2010). *Testing a conception of how school leadership influences learning*. Unpublished paper. Toronto, CA: U. of Toronto/OISE.

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

Maurer, R. (1996). *Beyond the wall of resistance.* Austin, TX: Bard Books.

McLaughlin, M. (1990). The Rand change agent study revisited. *Educational Researcher, 5,* 11-16.

Mortimore, P. (1993). School effectiveness and the management of effective learning and teaching. *School Effectiveness and School Improvement, 4*, 290-310.

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

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

Murphy, J. & Meyers, C. (2008). *Turning around failing schools: Leadership lessons from the organizational sciences*. Thousand Oaks, CA: Corwin.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides and R. E. Schumacker (Eds.), New developments and techniques in structural equation modeling (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum.

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

Nettles, S., & 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, 82,* 724-736.

Nicolaidou, M., & Ainscow, M. (2005). Understanding failing schools: Perspectives from the inside. *School Effectiveness and School Improvement, 16*(3), 229-248.

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

Opdenakker, M., & Van Damme, J. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal, 33*(2), 179-206.

O'Toole, J. (1995). *Leading change.* San Francisco: Jossey Bass.

Ouston, J. (1999). School effectiveness and school improvement: Critique of a movement. In T. Bush, L. Bell, R. Bolam, R. Glatter, and P. Ribbins (Eds.), *Educational management: Redefining theory, policy and practice* (pp. 166-177). London: Paul Chapman.

Purkey, S., & Smith, M. (1983). Effective schools: A review. *Elementary School Journal, 83*(4), 426-452.

Purkey, S., & Smith, M. (1985). School reform: The district policy implications of the effective schools literature. *Elementary School Journal, 85*(3), 352-389.

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

Reynolds, D., Teddlie, C., Hopkins, D., & Stringfield, S. (2000). Linking school effectiveness and school improvement. In C. Teddlie and D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 206-231). London: Falmer Press.

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

Rogers, E. (2003). *Diffusion of innovations*. New York: Simon and Schuster.

Saphier, J., & King, M. (1985). Good seeds grow in strong cultures. *Educational Leadership, 42*(6), 67-74.

Sarason, S. (1982). *The culture of the school and the problem of change* (Revised Edition). Boston: Allyn & Bacon.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: CRC Press.

Schein, E. (1996). Culture: The missing concept in organization studies. *Administrative Science Quarterly, 41*, 229-240.

Schön, D. (1983) *The reflective practitioner,* New York: Basic Books.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Doubleday.

Silins, H., & Mulford, B. (in press). Re-conceptualising school principalship that improves student outcomes. *Journal of Educational Leadership, Policy and Practice.*

Sleegers, P., Geijsel, F., & Van den Berg, R. (2002). Conditions fostering educational change. In K. Leithwood et al. (Eds.), *Second international handbook of educational leadership and administration* (pp. 75-102). Dordrecht: Netherlands: Kluwer Academic.

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

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

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer.

Teddlie, C., Stringfield, S., & Reynolds, D. (2000). Context issues within school effectiveness research. In C. Teddlie and D. Reynolds (Eds.), *International handbook of school effectiveness research* (pp. 160-183). London: Routledge.

van den Berg, R., & Ros, A. (1999). The permanent importance of the subjective reality of teachers during educational innovation: A concerns-based approach. *American Educational Research Journal, 36*(4), 879-906.

Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*, 1-16.

Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *34*(3), 398-425.

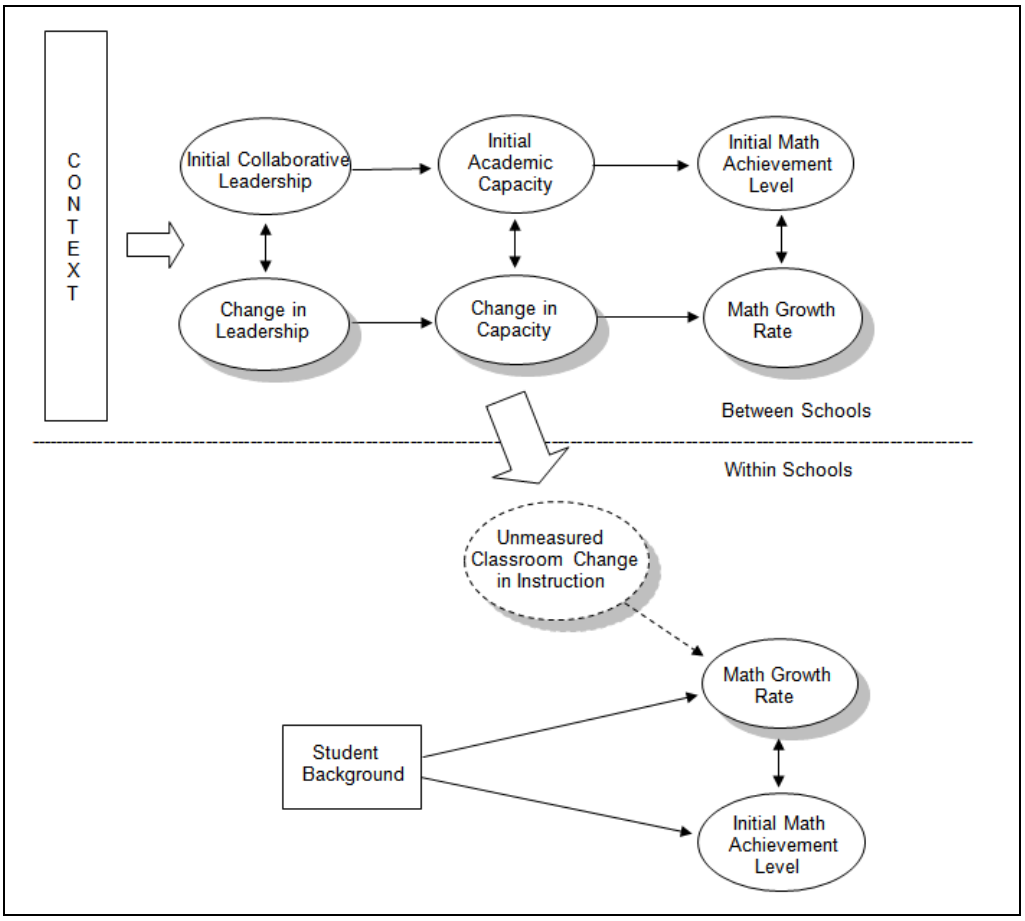

Figure 1. Conceptual Model of School Improvement Leadership and Student Learning

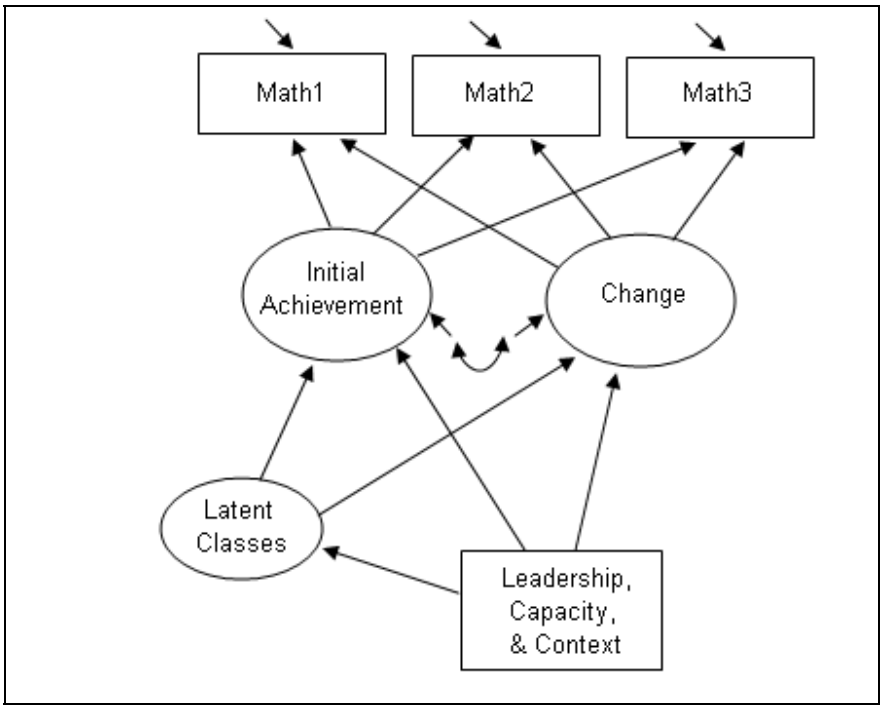

Figure 2. Proposed School-Level Latent Class Model

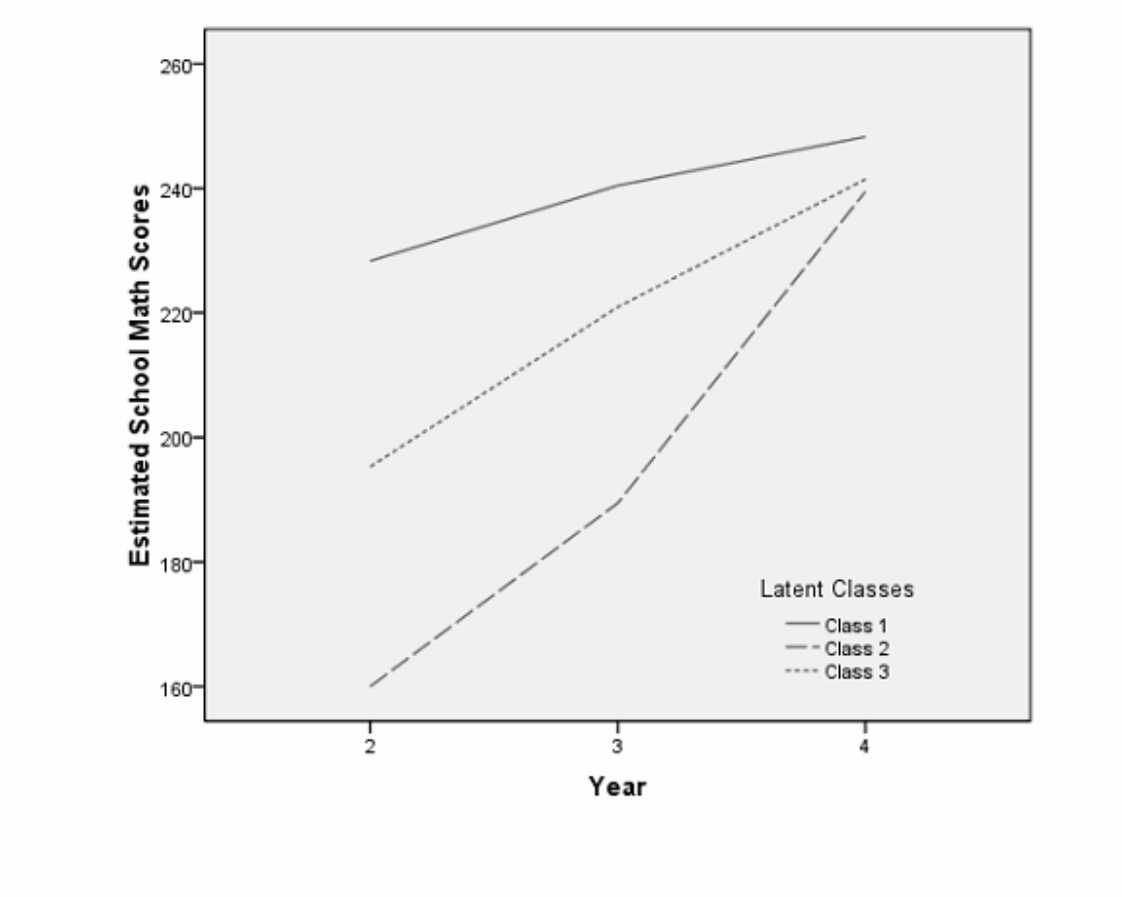

Figure 3. Latent Class Math Growth Trajectories

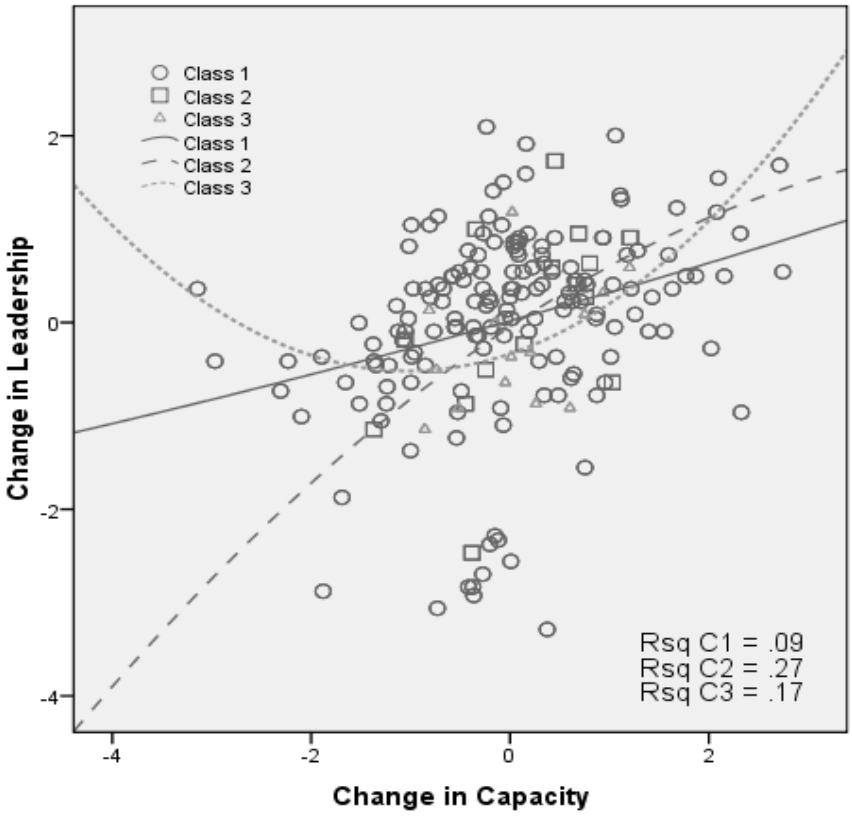

Figure 4. The Contribution of Leadership and Academic Capacity for Different Classes of Improving Schools

Table 1. Descriptive Statistics for Between-School Variables in the Model (N = 193)

VARIABLE NAME	MEAN	SD	MINIMUM	MAXIMUM
Context				
Enrollment	468.98	212.99	62.00	1278.00
ELL (%)	8.45	9.02	0.00	61.00
Low SES (%)	50.49	22.63	0.00	97.00
Underrepresented Mean (%)	51.16	23.97	3.00	97.00
School Composition	-0.03	0.95	-1.98	2.31
Staffing				
Same Principal	0.31	na	0.00	1.00
Staff Stability (%)	57.28	14.13	9.52	93.33
Met Licensing Criteria (%)	84.05	16.29	12.10	100.00
School Achievement				
Initial Math Level	247.45	35.86	108.33	294.09
Math Growth Rate	16.52	15.51	-11.91	76.52
Initial Distributed Leadership				
Leadership (%)	75.01	11.52	33.03	98.59
Leadership Factor Scores				
Year 1	0.00	0.12	-0.43	0.23
Year 3	0.03	0.14	-0.45	0.26
Year 4	0.02	0.14	-0.44	0.25
Initial School Improvement Capacity (%)				
Learning Standards (%)	87.10	6.32	69.14	98.67
Student Support (%)	78.48	10.93	37.63	98.77
Capacity (%)	74.53	11.82	40.01	99.11
Communication (%)	81.16	10.78	32.56	98.48
Focused School Improvement (%)	78.41	11.26	47.22	97.35
Involvement (%)	83.31	9.89	39.06	98.24
Safety/Well Being (%)	84.91	10.26	46.51	98.91
Improvement Capacity Factor Scores				
Year 1	0.00	0.24	-0.70	0.24
Year 3	0.07	0.23	-0.40	0.26
Year 4	0.09	0.24	-0.47	0.28
Student Background				
Low SES	0.45	na	0.00	1.00
English Services	0.07	na	0.00	1.00
Special Education	0.11	na	0.00	1.00
Female	0.49	na	0.00	1.00
Minority	0.50	na	0.00	1.00
Changed Schools	0.16	na	0.00	1.00

Table 2. Standardized Discriminant Function Coefficients

Variable	Function 1	Function 2
School composition	0.89	0.39
Improvement capacity	0.77	-0.66
Collaborative leadership	-0.44	0.67
Average teacher quality	0.31	0.44
Teaching staff stability	0.06	0.22
Same principal	0.04	0.28

Notes

[i] Our initial examination of the measurement model used to define the latent improvement capacity and leadership factors at three points in time suggested that school improvement capacity and leadership displayed measurement invariance over time (defined as invariant factor structure, item loadings, and item intercepts). We then provided evidence that teachers' perceptions about schools' underlying capacity and collaborative leadership changed over time (as summarized in Table 1).

[ii] One possible reason for the lack of association between school context and teacher mobility is that mobility is driven primarily by retirement patterns and the union contract specifying transfer and hiring procedures (based on seniority) rather than by principal discretion in hiring teachers.

[iii] We found only low SES status and math scores were significantly related to student likelihood to change schools or enter the cohort late. In each case, the identified relationships were in a different direction for each dichotomous outcome. At the school level, there was no consistent predictor identified across both dichotomous outcomes [i.e., staff stability was negatively related to student likelihood to change schools; school composition (percentages of low SES, ELL, and SPED students) was positively related to students' likelihood to enter the study late].