Educational Administration Quarterly
Vol. 36, No. 4 (October 2000) 513-552

Examining the Impact of School Quality on School Outcomes and Improvement: A Value-Added Approach

Ronald H. Heck

There has been a recent state-level emphasis on monitoring student outcomes to develop comprehensive school accountability. Such monitoring systems can include content standards and benchmarks to measure progress, statewide assessment instruments, and school report card data that policy makers, school personnel, and parents can also use to compare schools. To be fair, school comparisons should somehow take into consideration differences in communities and the background characteristics of students who live in these communities. Little previous research has examined what report card information might be used to identify school indicators that are related to student achievement and improvement gains across school settings with diverse student composition. The purpose of this study is to present an approach to statewide school comparison that focuses on the value-added effects of report card indicators of elementary schools' educational environments on school achievement and school improvement after making school-level adjustments for student differences.

Current state legislative policies across the United States aimed at strengthening educational accountability through standards-based practice, parent choice, and charter schools emphasize policy-makers' beliefs that schools may be evaluated in terms of their effectiveness in educating their students. School personnel are being impelled to change the status quo of professional practice in various ways with the expectation that they will be accountable for improving student performance. School effectiveness research, in part, has been a driving force behind such efforts, determining

Author's Note: The author is grateful to Vicki Rosser, Scott Thomas, Susan Wiley, Liang Zhang, the editors of the journal, and three anonymous reviewers for helpful comments on earlier drafts of the manuscript.

that school structure and the quality of educational processes (e.g., leadership, values and expectations, climate, teaching practices) can make a difference in student achievement (Brookover & Lezotte, 1977; Creemers, 1994; Edmonds, 1979; Hallinger & Heck, 1996; Mortimore, 1993; Reynolds & Packer, 1992; Witte & Walsh, 1990).

At the state level, the recent emphasis on monitoring student outcomes has been to develop comprehensive school accountability systems. These systems include curriculum content standards for what students should know, benchmarks to measure progress, statewide assessment instruments, and school report card data that policy makers, school personnel, and parents can also use in making comparisons among schools. Currently, every state except Iowa, which has voluntary administration of a basic skills test, requires at least one form of a statewide test (Olson, 1999). Most of these states use some combination of test scores and a variety of other school context data from their report cards to evaluate school performance.

School report card information often includes student socioeconomic status (SES), attendance and graduation rates, course-taking patterns, safety, teacher qualifications and salary, class size, parent involvement, and parent satisfaction. Thirty-six states have developed some type of school report card (Jerald & Boser, 1999), with another four to report this information in 2000 (Olson, 1999). More than half of these states assign performance ratings to their schools or at a minimum, identify low-performing schools, using some type of criteria (Jerald & Boser, 1999). Another eight states have recently passed legislation that will allow them to use these evaluations in the future (Viadero, 1999). The information provided across states, however, is not consistent in content or format and is used in widely different manners. Although the stakes are high, including sanctions or rewards for performance, to date there is little research on the relationship of these data to school outcomes. Moreover, it remains unclear what information schools can actually use to create improvement.

Although parents generally view high test scores as an indicator of a good school, educational practitioners have often been reluctant to rely solely on these types of indicators of educational quality (Salganik, 1994). One important issue is that the use of student outcomes as an indicator of educational quality raises concerns about test fairness (Oakes, 1989; Salganik, 1994). Because of differences in schools in terms of socioeconomic conditions (e.g., poor inner-city and wealthy suburban schools), parent background, staff characteristics, and student composition, as Salganik argues, most would agree that the effort necessary to produce outcomes is different. These differences are often not recognized, however, when comparing student performance across schools. In the past, newspapers generally published tables of schools' raw examination results as indicators of educational quality in an effort to improve accountability (Sammons, Nuttall, Cuttance, & Thomas, 1995).

To be fair, comparisons among schools for accountability purposes should somehow take into consideration differences in communities and the background characteristics of students who live in these communities (Darling-Hammond, 1994; Sammons et al., 1995). When student composition is ignored, it becomes more difficult to disentangle what value the school contributes to students' educational development. Analyses of school effects on student achievement that do not control for the impact of sociocultural factors, in addition to prior attainment, are likely to favor schools with more advantaged contextual factors, thus making schools with disadvantaged factors appear less effective than they are in reality (Willms & Kerckhoff, 1995). In an effort to avoid penalizing schools for problems beyond their control, a few states (e.g., Kansas, Kentucky, North Carolina, Tennessee) base their school or district accountability systems on improvements that students make. By comparing students' results with their past performance, the effects of their differing backgrounds are diminished, and the focus is placed more squarely on the effects of schooling. Another approach adopted by a small number of states is to adjust school performance expectations by factoring in students' cognitive ability (Indiana) or other student demographic variables, as in Kansas and New Mexico (Jerald & Boser, 1999).

The purpose of this study is to present an approach to statewide school comparison that focuses on the value-added effects of elementary schools' educational environments (e.g., administrative leadership, expectations for students, instructional processes, climate) on school achievement and school improvement. These school indicators are examined after making within-school adjustments for differences in student composition. This statistical adjustment ensures that the estimates of school performance are not biased against those schools with more challenging contexts. Although state report card systems encourage the collection of various data on schools, this information is seldom used in a manner that might identify what types of school indicators contribute to gains in student achievement across school settings that are widely different.

OVERVIEW OF THE STUDY

Strategies for Comparing School Outcomes

There are several different strategies for comparing schools for accountability purposes. A first approach emphasizes *gross productivity*, or the average performance of students in a school or district (Willms & Kerckhoff, 1995). Although this approach to school comparison has been most often used because it indicates actual levels of performance, it provides a biased view of what schools contribute to student learning because of the failure to consider student composition variables and previous learning levels. More specifically, this approach incorrectly assumes that all of the observed variability in raw scores is due to differences among the schools. Currently, most states use this approach, however, typically ranking schools by their raw score outcomes and providing report card data as background information about schools and students to help policy makers, school personnel, and parents interpret the observed outcomes (Salganik, 1994). No attempt is made, however, to link the report card information directly to the outcomes produced.

A second approach is to compare schools against others in a comparison group that are similar in terms of selected contextual and student background factors. Currently, nine states (e.g., Maryland, New Jersey, Texas) include information on their report cards that allows comparisons of a school's test scores with scores of similar schools or districts (Jerald & Boser, 1999). Although this approach attempts a more equitable comparison of schools by matching them in terms of some key criteria, a disadvantage is that it often relies on arbitrary cutpoints within the data to form the comparison groups (Salganik, 1994). Moreover, in this approach, the formation of comparison groups often uses statistical methods that do not take into account either the full range of schools in the data set or the multilevel structure (students nested within schools) of the data (Willms & Kerckhoff, 1995).

A third approach is to adjust the academic outcomes statistically for key indicators known to affect student achievement before making school comparisons. From a policy standpoint, this is an attractive means of determining how much value a school adds to student learning, given its particular student challenges. The assumption underlying a value-added approach is that students' achievements are significantly affected by their backgrounds and other contextual conditions (community factors). Multiple regression is used to develop a prediction that represents the best estimate of the achievement outcome, conditional on the selected, theoretically relevant, contextual information included in the model (Bosker, Kremers, & Lugthart, 1990; Goldstein, 1987; Hill & Rowe, 1996; Salganik, 1994; Sammons, Nuttall, & Cuttance, 1993; Scheerens, 1993; Willms & Kerckhoff, 1995). The approach emphasizes "net productivity," or average school achievement, after adjustments for the contextual indicators. The educational value that the school adds is expressed by a regression residual, or the difference between the observed achievement score and the score that would be predicted from the contextual information (Hill & Rowe, 1996; Nuttall, Goldstein, Prosser, & Rasbash, 1989; Sammons et al., 1995). Schools in which the observed scores are higher than the predicted scores are considered to be adding value to their students' learning.

Despite the general interest in this approach, however, it has not been frequently used in making school comparisons, in part because of the complexity of presenting the results to policy makers and parents. For example, of those 20 states currently rating their schools' performances in some manner, only three (Alabama, Indiana, Rhode Island) use outcome measures predicted by some set of student demographics as part of their school rating system (Jerald & Boser, 1999). Critics of this approach argue that the absolute level of school performance is obscured because the analysis focuses on the adjusted school achievement scores (residuals), as opposed to the raw achievement scores. More specifically, lower observed test scores are statistically adjusted upward for schools with more challenging student demographics. Of course, this type of analysis will change the numerical ranking of schools by raw achievement scores alone. Moreover, because scores are statistically adjusted against a norm group (e.g., other schools in the state or district), some schools will emerge as winners and others will be losers.

A Multilevel Approach to School Comparison

In this study, a fourth approach to school comparison is presented—an approach that acknowledges the special multilevel features of schools as organizations. Because a group of students resides in a common community and school, the students are likely to share similarities, including common background characteristics, experiences, and values. Moreover, they are assigned to particular classes and teachers within the school. These various school structures create common educational experiences among groups of students. Ignoring such nested data structures can lead to false inferences about the relations among variables as well as missed insights about the processes one is studying (Heck & Thomas, 2000).

In the past, researchers have had considerable conceptual and methodological difficulty analyzing models in which individuals are nested within a series of multilevel socio-organizational groups. For example, when data are analyzed using students as the unit of analysis, the possibility is removed of disentangling student effects on learning from school-level effects on learning (Seltzer, 1995). In contrast, when the school is the unit of analysis, all of the variability due to students within the school is reduced to a single school achievement score. This tends to overemphasize differences between schools, because most of the variability in student outcomes is due to differences among individual students within each school (Hill & Rowe, 1996). Standard statistical tests depend heavily on the assumption of independence of observations that accompanies simple random sampling. Because nested data produce pockets of similarity among the individuals comprising each group, the standard errors of parameters in the model are underestimated—potentially resulting in greater likelihood of the false attribution of statistical effects where none might exist (Burstein, 1980; Raudenbush & Bryk, 1986).

Over the past decade, there has been a substantial increase in the use of multilevel regression (or random coefficients) modeling in school-effects research (e.g., Goldstein, 1987; Hill & Rowe, 1996; Lee & Bryk, 1989; McDonnell, 1995; Muthén et al., 1995; Raudenbush & Bryk, 1986; Salganik, 1994; Sammons et al., 1993; Sammons et al., 1995; Scheerens, 1993; Wang, 1998; Willms & Kerckhoff, 1995). Multilevel modeling presents several advantages in comparing schools. First, it allows the researcher to investigate the extent to which clustering effects are present in the data. For example, the variation in achievement can be partitioned into individual-level and group-level components. Intraclass correlations, which describe the proportion of the total variance in an outcome that can be attributable to clustering effects, for achievement outcomes have been reported in multilevel school effects studies as somewhere between 10% and 20% of the total variance (e.g., Hill & Rowe, 1996; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Reynolds & Cuttance, 1992; Scheerens, Vermeulen, & Pelgrum, 1989). The amount of variability attributed to schools has been found to depend on such factors as the number of levels in the analysis (e.g., including classrooms as an intermediate level), the sampling strategy, the school level, and the type of academic outcome examined (Mortimore et al., 1988; Sammons et al., 1995; Seltzer, Frank, & Bryk, 1994).

Second, multilevel modeling provides a framework in which researchers can place explanatory variables at their correct level of the data hierarchy. This allows researchers to avoid aggregating or disaggregating the data to a single level of analysis. In multilevel modeling situations, at least two submodels are specified, one for the data at the individual-student level, and another for explaining the unknown distributions for each random individual parameter at the school level. More specifically, student composition variables can be incorporated into a student-level regression equation developed for each school that adjusts the school means for student differences. This has the effect of equalizing the individual-level characteristics across the schools. The school-level model consists of variables (e.g., contextual characteristics, school processes) that might account for variation in the adjusted school means.

Third, multilevel regression modeling also allows the researcher to investigate the variation in regression coefficients (slopes) comprising each school's regression equation across the set of schools. Confining ourselves to the analysis of school means can hide important differences concerning the distribution of academic outcomes within schools over time; that is, to what extent do students improve their academic standing in some schools? When we fix the student-level predictors (like achievement effects), we assume that the effects are the same across schools. In contrast, when we allow slope coefficients to vary randomly across schools, we assume that the effects are distributed differently across the sample of schools. Each slope, therefore, has a mean (or average effect) and a variance.

Implicit in this type of model specification is that schools differ in the manner in which the long-term learning effects may vary for students with different levels of initial educational attainment and backgrounds (Sammons et al., 1995). By examining the distribution of regression slopes describing the effect of previous learning on current learning across schools, we can identify schools that produce greater or lesser impact on student academic improvement. As Willms and Kerckhoff (1995) suggest, this can be seen as an analysis of inequities (or value added)—that is, the extent to which schools vary in producing learning gains, given certain levels of student entry variables (e.g., background, prior achievement).

To produce this type of analysis, test scores must be collected on multiple occasions for students, and a slope capturing the rate of progress over time calculated for the students within each school. At a minimum, at least two measurements on each individual are required. It should be noted, however, that this provides only minimal information on individual change and requires the assumption that the growth, or improvement, is linear over time (Willett & Sayer, 1996). For example, flatter slopes indicate less relationship (or change) between initial or previous achievement and current achievement levels, whereas steeper slopes indicate greater change. Moreover, slope parameters in the individual-level model can be adjusted for student composition variables, although few previous studies have investigated variation in school effects on learning over time as measured by school-level slope residuals (Bosker & Scheerens, 1989; Gray & Simes, 1991; Raudenbush, 1989; Sammons et al., 1993; Sammons et al., 1995; Seltzer et al., 1994; Willms & Raudenbush, 1989). After these adjustments, positive slope residuals (i.e., in which the observed slope is greater than the slope predicted from the student composition variables) suggest schools in which students make greater learning gains over time. The variation in individual-level slopes can then be modeled as a function of school-level variables (e.g., contextual variables, school educational conditions) that account for these differences in improvement.

To conduct a value-added comparison of schools' achievement and improvement, several conditions must be met (Salganik, 1994).

(a) Individual-level data on students must be available.
(b) The student background characteristics included in the model must be related to student performance.
(c) The student characteristics must be beyond the control of the school; otherwise, the school could be expected to change them.
(d) The characteristics included should be accepted as legitimately associated with the educational challenges facing the school (e.g., poverty, cultural differences in achievement).

It is important to emphasize, however, that comparing achievement separately for students of different ethnicity, socioeconomic background, or educational status (e.g., special education, language background) can be very politically sensitive.

Developing the Proposed Value-Added Model

Any attempt to model the richness of organizational life must begin with some admissions of its limitations. Proposed theories often become problematic when they attempt to model the actual detail of real organizations, because organizations are socially constructed realities with complex sets of interrelationships among their internal and environmental processes (Hallinger & Heck, 1998). It is obviously a reduction in reality for researchers to believe they can capture the richness and complexity of school processes in a finite set of variables. Researchers who seek to develop valid theoretical models and apply appropriate analytic techniques to assess how those models work in the world confront a formidable set of tasks.

Concerns can therefore be raised about any attempt to quantify and measure the quality of central components of the school's educational processes (e.g., its goals, leadership, values, expectations, communication, decision making). Although this reduction of reality can be considered a limitation in researching organizations, there is usefulness to the approach if it begins to unlock the black box surrounding the various avenues through which school personnel and practices affect school outcomes (Hallinger & Heck, 1998). Defining and measuring such avenues and demonstrating their impact on outcomes is one means of establishing the construct validity of a proposed model of value-added school comparison.

The research on school effects over the past two decades has identified several sets of factors that can make a difference in students' learning

(Edmonds, 1979; Reynolds & Packer, 1992; Witte & Walsh, 1990). Although results have not always been consistent across individual studies, taken together, the variables identified make up a conceptual framework indicating the importance of a school's contextual conditions—its structure, policies, personnel, and processes and its students' backgrounds—in determining student achievement (Creemers, 1994; Hallinger & Murphy, 1986; Hallinger & Heck, 1998; Heck & Marcoulides, 1996; Leithwood, 1994; McDonnell, 1995; Mortimore, 1993; Reynolds & Packer, 1992; Sammons et al., 1995; Witte & Walsh, 1990).

Contextual conditions. Communities and their schools constitute a context for learning. Learning opportunities are often different for students in schools in which the SES and ethnic composition vary greatly (Wiley & Yoon, 1995). Researchers (Hallinger & Heck, 1998; Hallinger & Murphy, 1986) note that these factors appear to influence principal leadership and the shaping of school processes (e.g., mission, expectations, norms and values, class formulation, access to curriculum). Students from a lower SES are often at a disadvantage in gaining access to quality curricula and teaching (Guiton & Oakes, 1995). Studies have also suggested that minority students are more likely to be put into classrooms with less learning opportunities, even when ability is considered (e.g., Gross, 1993; Jackson, 1982; Raizen, 1993). The inclusion of contextual variables therefore facilitates making more refined achievement comparisons among schools because of their known effects on learning. Contextual measures that have been used include per capita income, percentage of students with family income below poverty status, percentage of students on free or reduced lunch, percentage of minority enrollment, percentage of adults who are high school graduates, and percentage of students with limited English proficiency (e.g., see Salganik, 1994; Sammons et al., 1995).

School variables. School variables are important to consider in comparisons of school outcomes because they represent information about how schools are organized and run, how resources are allocated, how classrooms are formed, and how students are taught (Edmonds, 1979; McDonnell, 1995; Mortimore, 1993; Reynolds & Packer, 1992). Such decisions affect students' opportunities to learn (Burns & Mason, 1998). How well the school staff and parents are able to organize and coordinate the work life of the school (e.g., its mission and goals, governance, curriculum and instructional techniques, student groupings) shapes not only the learning experiences and achievements of the students but also the environment in which this work is carried out (Heck, Larsen, & Marcoulides, 1990).

There are obviously many indicators of the quality of school conditions and instructional processes, and it would be a mistake to think of any set of indicators as complete (Heck & Marcoulides, 1996). State report card data include a variety of information concerning academics and achievement, student characteristics, and school conditions (e.g., class size, teacher qualifications and salary, attendance, climate, and safety). As part of their report cards, a few states like Hawai‘i include additional data on parent, student, and staff perceptions about educational conditions in their schools (Jerald & Boser, 1999; Olson, 1999). In this study, several indicators of the quality of school conditions were defined and their collective impact on school achievement and improvement was examined. These indicators were

- principal leadership,
- high expectations for student achievement,
- an emphasis on academics,
- frequent monitoring of student progress,
- positive school climate, and
- positive relationships between the school and parent community.

In various combinations, these variables have been found to be related to school outcomes in previous research (e.g., Creemers, 1994; Edmonds, 1979; Hallinger & Heck, 1996; Hill & Rowe, 1996; Mortimore et al., 1988; Reynolds & Packer, 1992).

The attractiveness of defining and measuring the quality of school conditions for policy research is that school personnel have considerable control over the school's processes—in sharp contrast to their lack of control over student and community characteristics (Heck & Marcoulides, 1996). These variables can provide information about the quality of students' educational environment (academic press, opportunity to learn) and help explain why their achievements vary across classrooms or schools (Barr & Dreeban, 1983; Hallinger & Heck, 1998; Hallinger & Murphy, 1986; Heck & Mayor, 1993; Hill & Rowe, 1996; Kaplan & Elliott, 1997; Lee & Bryk, 1989; Leithwood, 1994; Muthén et al., 1995; National Council on Educational Standards [NCES], 1992; Sammons et al., 1995; Wang, 1998). Although teacher, student, and parent reports on classroom and school processes are a valuable, efficient, and cost-effective source of information describing the quality of the school's educational environment, it can be difficult to obtain reliable and valid information and to formulate all of this information into school-level indicators of these processes (Burstein, 1992; Raizen, 1993).

Student background. Even though information about the quality of the school's educational environment has been valuable in evaluating what happens in classrooms and schools, its effect on student learning outcomes has been difficult to demonstrate (Wiley & Yoon, 1995). In part, this is because many other factors—home environment and SES—have equal or greater impact on student performance than school quality (Wiley & Yoon, 1995). In developing more equitable school comparisons, therefore, it is important to include student background variables to adjust within-school means for the presence of composition variables that influence achievement (Muthén et al., 1995; Sammons et al., 1995). Muthén and colleagues emphasize the need to control for prior performance. Student composition variables that have been found to affect achievement include

- prior achievement
- gender
- ethnicity
- SES
- language background
- special education status

(Hill & Rowe, 1996; Sammons et al., 1995).

In Figure 1, the proposed value-added model for school comparison is presented. This study attempts to answer the questions,

Research question 1: What characteristics of schools help explain why some schools have higher adjusted outcomes than others?

Research question 2: What characteristics of schools help explain why some schools have greater academic improvement than others?

The main goal of this analysis is to study the partial effects of the quality of school conditions on achievement and improvement after controlling for student-level and school-level contextual factors. This type of analysis corresponds more closely to what students know as a result of their educational experiences.

Individual-Level Model

For the individual-level model in Figure 1, student composition variables (age, gender, ethnicity, SES, and special education status) and previous third grade learning (total reading, total math, and total language) are proposed to

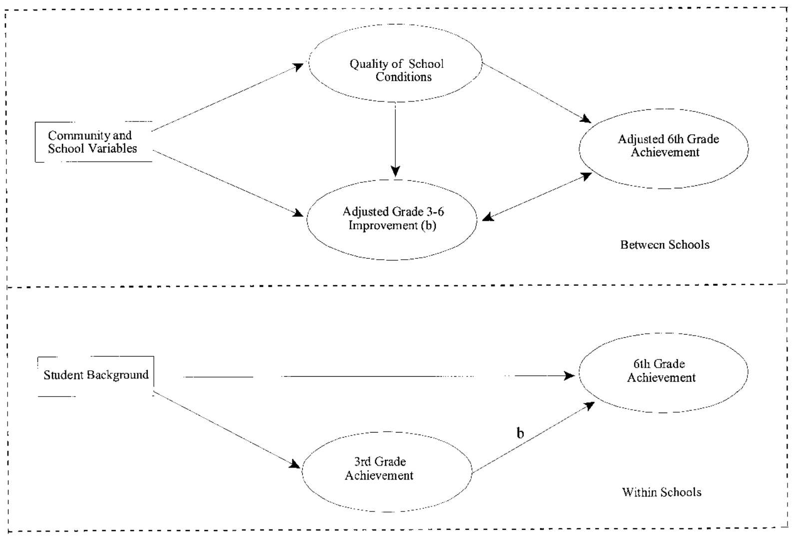

Figure 1: Proposed Within- and Between-School Models of Variables Affecting Student Improvement and Achievement
NOTE: b = average learning effect in each school, adjusted for student composition variables.

affect sixth-grade achievement (total reading, total math, and total language). Grand mean centering the predictors in the individual-level regression model yields an intercept that can be interpreted as the adjusted mean for each school (Bryk & Raudenbush, 1992). This effectively equalizes schools for differences among individuals on each student composition variable. Similarly, the slopes for previous learning represent the average Grade 3 to Grade 6 within-school improvement after adjustment for the student composition variables (represented by "b" in Figure 1). In the individual-level model, the school intercept for each sixth-grade outcome and the slope coefficients between each third-grade score and the corresponding sixth-grade score are defined as random parameters across the sample of schools.

School-Level Model

In the school-level model, contextual and school variables are regressed on the intercept and slope residuals from each school's individual-level regression equation. This means that we attempt to explain the different values across schools for the random regression coefficients by the set of school-level indicators (Hox, 1995). More specifically, in the school-level model, the intercept residuals are hypothesized to define a latent achievement factor with the effects of student composition removed. Similarly, the slope residuals are hypothesized to define a latent improvement factor, also equalized for the student composition variables. The achievement and improvement factors are proposed to vary as a function of a third latent factor that measures the quality of school conditions, a community SES variable, and several other school-level report card indicators (e.g., school size, teacher data, attendance, percentage of special education students). Equalizing schools in their student composition factors before making comparisons between schools should allow a more refined examination of how the quality of the school's educational environment (principal leadership, academic press, climate and culture, students' learning experiences, teaching practices, and curriculum coverage) affects school outcomes.

METHOD

Participants

Participants in this study were a subset of Hawai'i's public school students (approximately 188,000 students in 243 elementary and secondary schools). There are a number of differences to keep in mind about education in Hawai'i. Hawai'i is the only state in this country that has one centralized public school system. Considered as a district, it is in the top 10 in the United States in terms of the number of students served (NCES, 1997). There is tremendous diversity in the rural and urban nature of its schools as well as considerable variation in the size of its schools (ranging from fewer than 100 students to more than 2,000 students).

All of the elementary schools with K-6 configurations in the state were included in the study (*N* = 122). Because student progress is only measured at Grade 3 and Grade 6 during children's early years, 54 elementary schools that had K-5 configurations and one school that had a K-2 configuration were eliminated. To be included in the study, students had to have complete data for the third- and sixth-grade standardized tests and had to have attended the same elementary school during that time (*N* = 6,970). The number of students in the study within the 122 schools ranged from a low of six to a high of 260, with a mean of 62.6 students and a standard deviation of 41.8.[1]

For the most part, the backgrounds of students participating in the study were similar to studies conducted in other states. Forty-nine percent of the students were female. Special education students made up 11% of the sample (against a state average of 10% and a national average of 12.4%). Non-English-speaking students made up a small percentage of the sample (10%). Although no direct measure of individual students' SES was available, the state's federally funded, subsidized lunch program provides information that can serve as a proxy for low SES. In 1996, 37% of the state's K-6 students participated in this program. In this study, 39% of the students received free or reduced-cost lunch. To put these numbers in national perspective, 11% of the state's children ages 6 though 11 were classified as below the poverty level in 1995, somewhat lower than the national average of 18% (NCES, 1997).

There is one demographic difference to keep in mind, however, when interpreting and generalizing the results of this analysis. The ethnic composition in Hawai'i is different from other states in the country. Asian and Pacific Island students make up 69% of the state's population of students, whereas Caucasian students make up only 23%, and Hispanic and African American students together make up only 7.5%(NCES, 1997). Similar to studies examining the educational progress of minority students in other settings, some groups of Asian and Pacific Island students in Hawai'i (e.g., Filipinos, Hawaiians, Samoans) have been found to achieve below-state averages on standardized tests over time (Kamehameha Schools/Bishop Estate, 1993). These students comprise almost half of the state's public-school students and, similarly, about half of this study's student sample (18% Filipino, 28% Hawaiian and part Hawaiian, and 3% Samoan). Although the ethnic composition of students is the major difference between participants in this study and those in previous studies, this should not be viewed as a major limitation, because the goal is to examine the effects of school process indicators on outcomes after equalizing schools on the identified student composition factors.

Variables Included in the Model

Individual-Level Variables

Achievement outcomes. Total reading, math, and language scaled scores on the 1996 Stanford Achievement Test, edition 8, for individual students were used to form the school means for the academic outcomes. For the sample, the mean unadjusted sixth-grade scaled scores were 643 for total reading, 660 for total math, and 645 for total language (see Table 2). In contrast, the state averages (N = 12,759) for sixth-grade students were 634 for total reading, 647 for total math, and 624 for language. This suggests that as a group, children who stayed in the same elementary school between third and sixth grade had slightly higher test scores than the state's sixth-grade population (which also includes students who had changed schools or were attending middle schools during sixth grade).

Prior achievement. The corresponding Grade 3 total reading, math, and language test scores (1993 Stanford Achievement Test) were included for each student to define prior achievement. For ease of presentation, these scores were standardized ($M = 0$, $SD = 1$) with reference to the cohort of students who participated in the study.

Other student composition variables. Several other student composition variables were also included in the study. SES was defined by using free and reduced-cost lunch status as a proxy (coded 0 for the reference group and 1 = low SES). Ethnicity was coded 0 for the reference group and 1 for each ethnic group included. Gender (coded 0 = male and 1 = female), special education status (coded 0 = regular education, 1 = special education), home language background (0 = English, 1 = non-English speaking), and age (in months) were also included in the analysis.

School-Level Variables

Community SES. Census data from 1990 were combined to create a weighted factor score ($M = 0$, $SD = 1$) representing community socioeconomic status (comSES) using principal components analysis. The variables comprising the component (which accounted for more than 90% of the observed variance) were percentage living in poverty, percentage receiving public assistance, median income, percentage of high school graduates, per capita income, and the percentage of children in each school receiving free or reduced-cost lunch. Some variables were reverse coded before being combined into the final component such that higher scores indicate greater comSES.

Other school demographic variables. Several other school demographic variables were also investigated in preliminary models. Teacher experience and school attendance figures were found to be unrelated to the outcomes and were not retained in the final model. School size was a dichotomous variable, coded 0 (schools having fewer than 600 students) and 1 (schools having 600 or more students). The other school-level variable was percentage of special education students.

School conditions. The *Effective Schools Survey* (Hawai'i Department of Education, 1996) is administered at the site level to all certificated staff, all Grade-5 students, and a random sampling of parents (approximately 20% across grade levels within each school) on regular cycles (1994-1996, 1995-1997) in the state of Hawai'i. The purpose of the survey is to provide information from staff, students, and parents about school conditions as part of the state's report card data. The staff survey consists of 60 items measuring six indicators (10 items per indicator), whereas the student and parent surveys consist of 36 items (6 items measuring each indicator). All items are measured on Likert-type scales (1 = *strongly agree* to 5 = *strongly disagree*). For purposes of data analysis, each item was recoded into the percentage who agreed with the statement (e.g., "The principal makes student achievement the school's top goal.").

The specific school conditions that the surveys monitor include the quality of principal leadership (e.g., making student achievement a top goal, monitoring teachers' work, solving problems effectively, involving others in decision making to improve school curriculum), a schoolwide emphasis on academics (e.g., how teachers present work in class, how students participate in class, use of class time, student involvement in class), high expectations for student achievement (e.g., teacher beliefs about students, curricular emphasis on developing a wide range of skills, challenging academic work), frequent monitoring of student progress (e.g., teacher grading practices, homework, teacher management techniques), positive school climate (e.g., safe environment, clean and comfortable buildings, teachers demonstrate caring attitudes), and positive school-home relations (e.g., regular communication, parents feel welcome at school, school seeks parental involvement in decision making, parents are involved in school activities).

Survey response rates were calculated for staff (surveys returned divided by total staff size), parents (number of returns divided by number sampled), and students (fifth-grade surveys returned divided by fifth-grade class size) for a random sample of schools in the study.[2] The staff returns ranged from 67% to 100% (M = 89%, SD = 11%). Student returns ranged from 63% to 98% (M = 86%, SD = 9%). Parent return rates varied from 24% to 94% (M = 58%, SD = 21%). It should therefore be noted as a limitation that the parent perceptions included in the study are likely to be less representative of that group's views about their children's school conditions.

In compiling data from different sets of respondents to measure the quality of school educational conditions, it is important to acknowledge the conceptual aggregation that must take place. As Heck and Marcoulides (1996) suggest, it is difficult to translate theories of how school processes operate (e.g., leadership, values and beliefs, climate) into actual data that describe these processes. In this study, data were collected from many individuals, and there is obviously variability in how teachers, students, and parents in a particular school view its quality. The aggregation of responses to create measures of school conditions therefore represents a considerable reduction of reality.

To improve the psychometric properties of the variables, several steps were taken. First, a sizable number of survey items was used to define each school condition (i.e., six items for parents and students, 10 items for teachers). Therefore, six similar items for parents, students, and staff anchor each indicator (see appendix). Table 1 summarizes the Cronbach's alpha coefficients (a measure of each indicator's internal consistency across participants) for each group. Coefficients above .7 suggest considerable reliability for the indicator. The alpha coefficients ranged from .73 to .94, suggesting the reliability of measurement was relatively strong within each group of respondents. It should be noted that the reliability of measurement was greater for the set of staff indicators, because those school indicators were each composed of an additional four survey items.

Second, despite the differing numbers of items used to define the indicators, the correlations between groups on each indicator were found to be substantial. These correlations (not tabled) ranged from .5 to greater than .8 (mean correlation = .68, *SD* = .10). The strength of the correlations indicates considerable consistency in views about school conditions. Given the considerable correspondence of views, it seemed a reasonable compromise to aggregate the data across groups to produce one composite school mean for each indicator. This aggregation was necessary because of limitations in the number of model parameters that can be estimated with the number of schools in the study. The composite school indicators in Table 1 were all judged to be reliable (i.e., leadership = .80, academic emphasis = .86, high student achievement expectations = .84, frequent monitoring of student progress = .83, positive school climate = .89, and positive home-school relations = .90).

As a final step, a school-level factor score was estimated for each school indicator, using principal components analysis. Each school's resulting observed score for each indicator represents a weighted composite ($M = 0$, $SD = 1$) that accounts for the maximum variability observed in the correlations among the three role groups (Marcoulides & Hershberger, 1997). Because of the coding scheme, larger positive coefficients indicate stronger perceptions about the quality of school conditions and processes.

Data Analysis

The analysis of multilevel data presents a number of challenges. Multilevel models require us to extend single-level univariate and multivariate

TABLE 1
School Indicator Reliability

School Indicator	*Composite**	*Cronbach's α Coefficients*		
		Staff	*Parents*	*Students*
Principal instructional leadership	.80	.92	.90	.73
Emphasis on academics	.86	.93	.84	.78
High expectations for achievement	.84	.91	.84	.75
Frequent monitoring of student progress	.83	.94	.83	.77
Positive school climate	.89	.90	.86	.82
Positive home-school relations	.90	.92	.86	.82

*Weighted school variable that combines parent, student, and staff perceptions.

analyses to more complex models with variables measured at different levels of analysis. Several different approaches for modeling multilevel data have been presented (e.g., see Bryk & Raudenbush, 1992; Hox, 1995; Muthén & Muthén, 1998; Raudenbush & Sampson, 1999), and corresponding software is becoming more readily available. The choice of an analytic paradigm requires the consideration of the research questions, the theoretical model to be investigated, the structure of the data, and the strengths and limitations of the various techniques and software programs.

To date, multilevel regression has been most often used to model the direct effects of predictors measured at several different levels of a data hierarchy on latent or observed outcomes (see Bryk & Raudenbush, 1992; Raudenbush & Sampson, 1999). Multilevel regression software programs like HLM (Bryk, Raudenbush, & Congdon, 1996) readily allow the modeling of variation in individual-level intercepts and slopes as well as the capability to include residuals in the analysis. To date, however, multilevel regression modeling has been only rarely applied to the estimation of indirect (or mediated) effects between variables (see Raudenbush & Sampson, 1999).

A second approach to multilevel modeling attempts to incorporate multilevel data into the general framework for analyzing mean and covariance structures with structural equation modeling (SEM). In the past, SEM has been a popular methodological choice because of its flexibility in representing a wide variety of theoretical models, including those with latent variables (that allows the incorporation of measurement error into the model), direct and indirect effects, and reciprocal causation. Applications of multilevel SEM models using a simultaneous estimation approach, however, have been rare in the literature because of difficulties in obtaining software that can be used to provide accurate estimations of model parameters. Although methodological work on this approach to modeling multilevel data is expanding rapidly, it is currently difficult to extend the approach beyond two levels or to model random slopes in the same manner as in the multilevel regression approach. More important for this proposed analysis, multilevel modeling with SEM yields individual-level estimates that are equivalent to group-mean centering in multilevel regression. Unlike grand-mean centering, which results in adjusted school intercepts, group-mean centering produces intercepts that are unadjusted for individual-level variables.

After a consideration of these issues, the data were analyzed using a two-stage approach to model estimation (Chou, Bentler, & Pentz, 1998; de Leeuw & Kreft, 1986; Hox, 1995; Tate & Wongbundhit, 1983). In this approach, one set of estimates is used for the individual-level model and a second set is used for the group-level model.[3] To conduct the analysis, HLM (Bryk et al., 1996) was first used to estimate the individual-level model for each school to adjust the random coefficients (school intercepts and improvement slopes) for students' background characteristics. Separate equations were estimated for reading, math, and language outcomes.[4]

From these individual-level models, the Bayesian estimates of the intercept and slope residuals (the observed scores minus predicted scores for each intercept and slope parameter) were saved in a school-level data file. Because Bayesian estimates take into consideration the precision of each school's regression equation, they provide more reliable predictions across the sampling distribution of schools in the study.[5] At the second stage of the analysis, the effects of the school-context and school-quality variables on the adjusted slopes and intercepts were estimated using structural equation modeling and LISREL 8.3 with maximum likelihood estimation (Jöreskog & Sörbom, 1999).[6]

RESULTS

The first step in a multilevel analysis is often to determine the percentage of variation in outcomes that lies within and between levels. Larger intraclass correlations, (which describe the percentage of between-group variation), suggest greater correspondence among group members.[7] Before adjustments for student composition, the sixth-grade intraclass correlations were .194 for reading, .231 for math, and .190 for language (see Table 2). For the same students' third-grade scores, the intraclass correlations were .210 for reading, .201 for math, and .179 for language (not tabled). The set of intraclass

TABLE 2
Descriptive Statistics of School Outcomes

Measure	M	SD	*Minimum*	*Maximum*
Unadjusted school outcomes*				
Read6	643.2	15.0	602.4	678.6
Math6	660.0	17.3	626.4	713.6
Lang6	645.2	12.9	613.1	679.9
Intercept residuals				
EBRead6	000.0	4.7	–9.9	15.5
EBMath6	000.0	8.9	–27.6	33.7
EBLang6	000.0	4.8	–11.8	11.0
Slope residuals				
EBRead3-read6	000.0	1.3	–2.9	4.1
EBMath3-math6	000.0	2.9	–8.9	7.4
EBLang3-lang6	000.0	1.9	–10.3	4.3

NOTE: *6th-grade intraclass correlations are .194, Read(ing)6; .231, Math6; and .190, Lang(uage)6. EB = empirical Bayes residuals.

correlations indicates that there is considerable between-schools variance in learning outcomes, consistent with the 10% to 20% between-school variance found in previous studies (e.g., Hill & Rowe, 1996). Some of this between-school variation would likely be reduced, however, if the intermediate nesting effect of classrooms within schools were considered (Hill & Rowe, 1996).

Individual-Level Analyses

In Table 2, the unadjusted total reading, total math, and total language means for each school are presented. The table also includes the range of empirical Bayes residuals in each subject area. Empirical Bayes residuals provide a more restricted, and therefore more reliable, range of school estimates than those calculated using ordinary least squares regression (Bryk & Raudenbush, 1992). The residuals ranged from –9.9 to 15.5 in total reading, from –27.6 to 33.7 in total math, and from –11.8 to 11.0 in total language. These ranges indicate the number of scaled score points that the schools are adding (or subtracting) to their students' expected academic outcomes (i.e., with 0.0 being the point at which the observed and expected achievement scores are the same).

The range of the residuals suggests that there is greater variability in school math residuals than in school reading or language residuals. To summarize the variability across the schools further, 57% were achieving below their expected score (0) in reading; 51% were below in math; and 52.5% were below in language. This suggests that somewhat more schools in the sample were doing worse than expected given their student composition factors, especially in reading.

The slope residual estimates in Table 2 ranged from –2.9 to 4.1 in reading, –8.9 to 7.4 in math, and –10.3 to 4.3 in language. This describes the relative steepness or flatness of the school achievement change between Grades 3 and 6. Similar to the analysis of intercept residuals, the mean slope residual across the sample of schools also corresponds to 0 (i.e., where the observed and expected slopes are exactly the same). To explain the meaning of these coefficients in more detail, we will look at the language residuals. In the lowest school, the Grade 3-6 achievement slope (–10.3) was flatter by about 10 scaled-score points than would be expected given its students' background characteristics. Moreover, this school would be about five standard deviations ($SD = 1.9$) away from the mean slope (0.0) for the set of schools. Table 2 also suggests that there was considerably less variation in reading slope residuals than in math and language slope residuals.

Table 3 provides the unstandardized coefficients for the individual-level models produced with HLM and centered on the grand mean. As shown in the table, the student background variables were almost all significant across the three outcome measures, which suggests that the sampled schools differed in terms of student composition variables.

After adjustment for the student characteristics and prior achievement, the intraclass correlations were further reduced in relation to the total variance. These reductions were particularly strong in reading (from .194 to .081) and language (from .190 to .084). This suggests that a substantial portion of the observed differences in reading and language between schools can be attributed to differences among students in those schools. This shrinkage of school-level variance can be seen in the generally tighter clustering of school intercept residuals for reading and language summarized in Table 2. In contrast, the intraclass correlation for math was less affected by adding the student-level variables (reduced from .231 to .195). We can conclude that student background factors do not affect between-school variability in mathematics nearly as much as they affect reading and language between-school variability.

Table 3 also indicates that the addition of the student composition variables accounted for substantial variance in achievement outcomes, both at the individual level (Level 1) and the school level (Level 2). For example, at Level 2, the student composition variables accounted for between 69%

TABLE 3
Within-School Student Composition Models

Fixed Effect	*Reading Coefficient*	*Math Coefficient*	*Language Coefficient*
Adjusted school mean	642.768683*	659.590903*	644.565240*
Female	–0.116282	1.761019*	4.225426*
Age	0.247229*	0.353424*	0.175816*
Low SES	–4.439431*	–4.269853*	–3.284686*
Previous achievement	26.840195*	28.471845*	20.278071*
Sped	–1.810867*	–3.412744*	–5.986660*
Filipino	–5.285413*	–2.154884*	–2.859613*
Hawaiian	–5.224724*	–3.471485*	–4.306228*
Samoan	–8.631834*	–2.567293*	–6.320327*
Intraclass correlation adjusted for Level-1 variables	.081	.195	.084
Intercept reliability	.750	.870	.750
3rd-6th grade achievement slope reliability	.310	.530	.460
Proportion of variance accounted for at Level 1	.661	.630	.545
Proportion of variance accounted for at Level 2	.858	.688	.800

NOTE: SES = socioeconomic status; sped = special education.
*$p < .05$.

(math) and 86% (reading) of the variance in the achievement outcomes. Finally, Table 3 also suggests that the intercept coefficients were measured quite accurately across groups, with average reliability coefficients ranging from .75 to .87. In contrast, the achievement slope coefficients were less reliably measured (with coefficients ranging from .31 to .46).

One preliminary test of the model's validity is whether ranking schools by their educational value added results in a different set of schools that are identified as outstanding. For illustrative purposes in comparing educational value added across the schools, the reading outcome and slope residuals are presented in Figure 2. For example, schools 122, 119, 121, and 120 were doing much better than expected in producing gains and outcomes, given their students' backgrounds and prior achievements. After the individual-level adjustments for student composition in reading, for example, only two of the top 10 schools (ranked by the size of the positive reading residuals) were located in high-SES communities. In contrast, without the adjustments

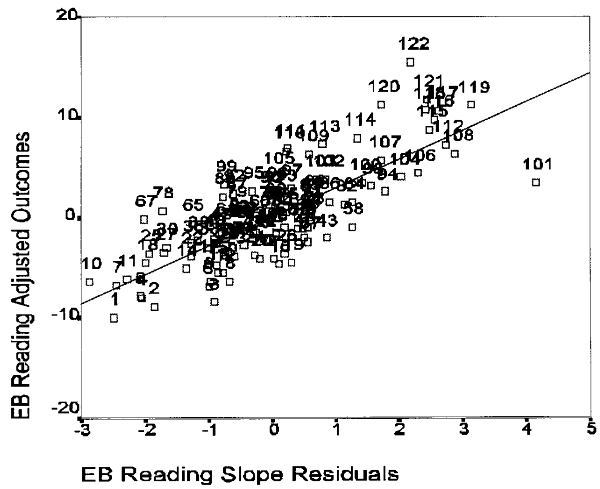

Figure 2: Reading Outcome and Improvement Residuals
NOTE: EB = empirical Bayes.

for student composition, the top 10 schools ranked by their sixth-grade reading scores were all located in high-SES communities.

Testing the Structural Model

Although Figure 2 shows an example of the variability across the schools, we do not know which school and community variables might account for this variability. The second part of the analysis focused on the school context and process variables that affect the adjusted outcomes and improvement. Because a theoretically driven model was proposed, the primary concern is with the fit of the model to the data. Without an adequate fit, it would be necessary to reconceptualize the model.

Several commonly used fit indices can be examined to determine the proposed model's fit to the data. The comparative fit index (CFI) compares a hypothesized model to a baseline, or null, model (defined as a complete absence of a covariance structure). The nonnormed fit index (NNFI) also measures the relative improvement in fit obtained by a proposed model compared to the null model (with a correction for the number of parameters in the model). With maximum likelihood estimation, both of these indices perform consistently across a variety of conditions including relatively small sample sizes (between 50 and 250) and moderate departures from normality (e.g., Hu

& Bentler, 1995). In general, acceptable values on these indices should be considerably above .9 for a good-fitting model. The root mean squared error of approximation (RMSEA) is another fit index that is widely used because it offers a close test of statistical fit for the model; that is, it allows for a discrepancy of fit per degree of freedom (Marcoulides & Hershberger, 1997). After this adjustment, it has the advantage of providing a test that gives a region for rejecting ill-fitting models on statistical grounds (i.e., where $p < .05$).

Given these guidelines, the structural model was determined to fit the data adequately (CFI = .97, NNFI =.96, RMSEA = .07, $p = .08$). This implies that the proposed model can be considered a plausible representation of the data. After the model fit is determined to be adequate, the individual parameters can be considered.

The model's parameters are displayed in Figure 3 and underscore several important results. First, the figure suggests that the observed variables were generally sufficient measures of the latent constructs. For example, the school quality factor was well measured by the six observed indicators, with loadings ranging from .59 to .97. Similarly, the achievement factor (consisting of the adjusted intercept residuals) was also well measured (math = .71, language = .74, reading = .81). In contrast, the improvement factor (consisting of the achievement slope residuals) was not as well measured (language = .23, math = .62, reading = .67). Correcting the model for measurement error, however, is an important step in obtaining more accurate information about the structural relationships between the latent variables.

Second, the structural parameters indicated a significant relationship between school quality and the achievement (.34) and improvement (.31) factors. This indicates that higher perceptions of the quality of the school's educational environment (i.e., principal leadership, academic press, high expectations, monitoring student progress, school climate, home-school relations) was related to better-than-expected academic outcomes at Grade 6, after controlling for the other school contextual variables. Higher school quality was also related to also better-than-expected learning improvement from third to sixth grades after similarly adjusting for the contextual variables.

Third, Figure 3 shows that school SES indicators had an impact on perceptions of school quality as well as on improvement and outcomes. More specifically, higher comSES was directly related to stronger perceptions of school educational conditions (.32), greater improvement (.40), and greater-than-expected outcomes (.53). There was also a small but significant, indirect (.11) effect of comSES on outcomes (through the mediating school-quality latent variable). This suggests that even after controls for within-school student factors, higher comSES is related to stronger school

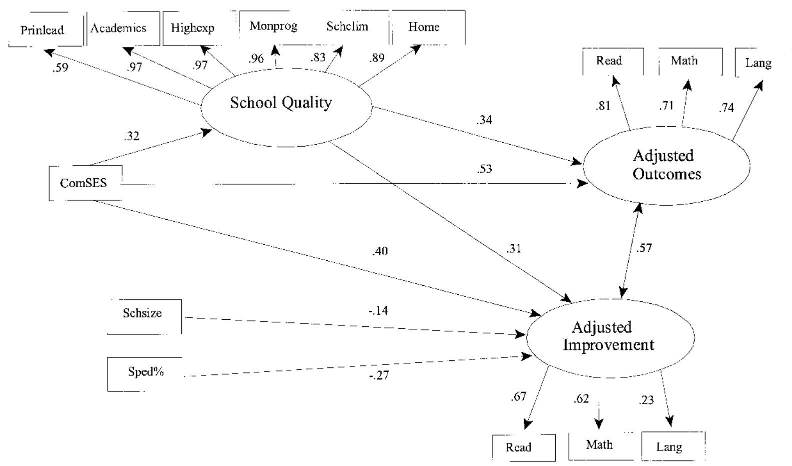

Figure 3: Structural Model of School Context and Quality Variables and Their Impact on Improvement and Achievement (significant standardized LISREL estimates, $p < .05$)

NOTE: Prinlead = principal's leadership; Highexp = high expectations; Monprog = frequent monitoring of student progress; Schclim = school climate; comSES = community socioeconomic status; Schsize = school size; Sped% = special education percentage.

quality and greater-than-expected learning. Students in schools educating children of wealthier families perform better than their counterparts in schools representing less affluent families.

Fourth, larger elementary schools produced smaller gains between third and sixth grades (–.14), and schools with higher percentages of special education students also produced smaller improvement gains between third and sixth grades (–.27). These variables were also negatively correlated with comSES, suggesting that larger schools and schools with greater percentages of special education students were found in less affluent communities.

Finally, the variables in the model accounted for about 30% of the variation in achievement residuals and about 35% of the variation in third- to sixth-grade school improvement (not tabled). Overall, the validation of the proposed model supports the view that the quality of the school's educational environment affects school improvement and school outcomes after the within-school adjustments for student composition have been made and the relevant between-school controls have been added. Moreover, the support of the proposed model also validates the constructs defined in the state survey on school quality indicators.

DISCUSSION AND POLICY IMPLICATIONS

As researchers have noted, nonschool factors such as student background, parent education, and community wealth have often been identified as important indicators of student achievement. Over time, we have seen an increase in single-parent families and an increase in poverty rates in many urban areas, as well as pervasive changes in social values. These types of societal trends are often beyond the control of the school. Ideally, schools should be held accountable to performance standards that reflect what they are contributing to students' achievement levels and growth; that is, we should focus on what school personnel contribute to children's learning given the realities in which they work (Darling-Hammond, 1994; Evans, 1999).

The purpose of this study was to demonstrate an approach to school comparison that focused on making preliminary adjustments for within-school student composition factors. Such statistical adjustments help ensure the equitable comparison of schools for accountability purposes. After making these adjustments, the intent was to explore how the quality of the school's educational conditions affects its achievement outcomes, as well as to explore the stability of these variables in explaining student improvement over time. The techniques used in this study underscore that policies and procedures to compare schools can give different impressions as to which schools are really adding value to children's learning.

Although almost all states mandate some type of state testing in connection with report-card data on school conditions, the type of information collected, the way in which it is compiled, and the use of this information vary greatly. Some states make comparisons based on average raw achievement outcomes; others compare the school's performance against a set benchmark (e.g., 70% of students passing the standard); whereas still others emphasize student improvement gains over time. A few also adjust for demographic factors such as poverty that are known to influence achievement outcomes. Comparisons between schools based on their educational value added go beyond typical comparisons in the media, which rank schools without regard to their various types of student and community differences.

Most important, the findings of this study demonstrate a pattern of achievement advantage favoring schools with stronger school educational environments. Schools rated as having higher quality educational environments created higher-than-expected outcomes after controlling for the composition of their students. This is encouraging news, because these types of effects have often been hard to produce in studies of school effects using standardized tests (Wiley & Yoon, 1995). These schools have principal leadership that is rated as more supportive and directed toward instructional excellence and school improvement. Teachers in these schools are given higher ratings for creating a classroom environment that emphasizes academics (e.g., using more class time for instruction, keeping students more frequently on task, providing students with extra help when needed). There are stronger expectations for student learning (e.g., providing challenging school work, informing parents and students of expectations). In these schools, the climate (e.g., safety, teacher caring) is seen in more positive terms. Finally, students, teachers, and parents perceive that the relationship between the school and home (e.g., communication, parent involvement, school responsiveness) is more positive.

Moreover, schools with stronger school environments also produced greater-than-expected improvement in student learning over time. Given these findings, it would appear necessary to continue to monitor the quality of schools' learning environments as a means of determining in which schools students enjoy greater opportunities to learn. School effects are by definition long term, cumulative, and apply to all students in the school (Hill & Rowe, 1996). For this reason, we should be interested in schools in which the quality of education (e.g., expectations, curriculum, teaching, monitoring of progress) is more uniform across classrooms and grades and the school leadership is more outstanding (e.g., greater parent support and involvement, more positive school climate).

It is also important to note the impact of the school's context on the outcomes it produces. Several contextual variables also exerted small-to-moderate effects on outcomes or improvements. In particular, community SES exerted moderate effects on both adjusted outcomes and improvement. School size and percentage of special education students also produced small negative effects on improvement. The results, therefore, confirm that many observed differences in student performance are due to characteristics of schools and their communities (Heck & Mayor, 1993; Witte & Walsh, 1990), even after adjustments for within-school composition effects. Even though many of these contextual variables are beyond the control of the school, it is important to understand and accommodate them because they influence the types of strategies and processes schools implement to improve learning outcomes (Hallinger & Heck, 1998; Hallinger & Murphy, 1986). For example, higher community SES was found to be associated with higher school quality (perceptions of leadership, academic environment, expectations, school climate, and home and school relations). The manner in which these relationships are fostered and maintained, however, remains elusive (Hallinger & Heck, 1996).

Continued efforts to develop fair and equitable models for school accountability should be encouraged. Reasonable evaluation of differences between schools, therefore, should emphasize appraisal that is context based; that is, any evaluation of the school must take into consideration its unique setting (student composition, school factors, community factors). From an evaluation perspective, conducting equitable performance assessments for school accountability purposes may include the need to consider student composition factors and contexts (e.g., through weighting, providing comparison groups) before making school comparisons. More specifically, some schools may face more imposing challenges, and this contextual situation has a measurable effect on performance. Schools should not be penalized for problems such as poverty that are beyond their control.

Political and practical cost dilemmas remain, however, for balancing equitable school comparisons and demands for high performance. Although state policy makers may wish to get all of their students up to a specific benchmark level, school improvement may require differing amounts of energy (e.g., staff expertise, staff development) and resources to accomplish for different types of students (Monk & Plecki, 1999). Moreover, measuring scores against a state average by developing an average regression equation can set a low benchmark if the state tends to perform near the bottom on national comparisons of progress. There are also practical costs involved in collecting this type of information, however, that will likely increase states' assessment budgets, and only a few states currently use this type of information directly in making comparisons of school outcomes. Collecting several longitudinal measurements on each student is desirable to provide more accurate assessments of student learning improvement, which also can increase assessment costs.

The goal of promoting equitable performance comparison, therefore, is not just to level the playing field for those in less advantaged schools, but also to promote the creation of a stronger sense of community among staff, students, and community as an essential part of equity (Kahne, 1994). This can translate into a commitment to improving the school's educational environment as a means for meeting its educational challenges and improving its outcomes. The results of this study imply that schools can be held accountable for the quality of educational processes that exist there provided that contextual factors such as student composition are controlled. Although most states already collect a variety of information as part of their school report-card system, there has been little prior research on the content and usefulness of this information in explaining how schools can improve student learning. This study demonstrates the relationship between the school's educational environment and its achievement and improvement. Despite the potential costs, focusing on educational value added is a promising approach to school comparison that is likely to provide further knowledge about how schooling affects student learning.

APPENDIX
EFFECTIVE SCHOOLS SURVEY

TEACHER/STAFF SURVEY	*STUDENTS*	*PARENTS*
Strong Instructional Leadership of the Principal	**Strong Instructional Leadership of the Principal**	**Strong Instructional Leadership of the Principal**
1. The principal makes student achievement the school's top goal.	1. The principal makes student achievement the school's top goal.	1. The principal makes student achievement the school's top goal.
2. The principal states the school's mission and goals in clear, concrete terms.	2. The principal tells students about the goals of the school.	2. The principal informs parents about the school's goals in clear, concrete terms.
3. The principal takes the lead to resolve instructional problems.	3. The principal and vice-principal handle trouble effectively.	3. The principal attempts to solve problems, not just talk about them.
4. School administrators work with teachers, students, and parents to develop the school's improvement plan.	4. Students help to make decisions about improving school courses and student activities.	4. The principal seeks parent and community input on how to improve the school.
5. There is ongoing two-way communication between the administrators and school personnel.	5. The principal often talks with students.	5. There is open two-way communication between the principal and parents.
6. The school administrators regularly observe classroom instruction.	6. The principal and vice-principal visit classrooms regularly.	
7. The school administrators regularly provide feedback to teachers with regard to their classroom instruction.		
8. Administrators and staff share in leadership roles, using individual and team strengths.		
9. The principal makes sure there are sufficient resources for effective instruction.		6. The principal makes sure there are enough instructional materials for students.
10. The principal ensures that there is an effective, ongoing system for evaluating the school's progress toward its goals.		

(continued)

APPENDIX Continued

TEACHER/STAFF SURVEY	*STUDENTS*	*PARENTS*
Strong Emphasis on Academics	**Strong Emphasis on Academics**	**Strong Emphasis on Academics**
11. Class time is used for instruction, not busy work.	7. In my classes, time is spent on learning, not busy work.	7. My child's class time is spent on learning, not busy work.
12. Teachers present academic work in interesting and varied ways.	8. Teachers present academic work in interesting and varied ways.	8. My child finds school work interesting.
13. Instruction is geared to having students actively involved in learning.	9. Students actively participate in classroom instruction.	9. In my child's classes, students actively participate in classroom instruction.
14. Students are given enough time to master the basic skills.	10. Students are given enough time to master the basic skills.	10. Students are given enough time to master the basic skills.
15. Students who need extra help get it.	11. Students who need extra help get it.	11. My child receives extra help when needed.
16. Teachers maximize student time-on-task.		
17. Teachers continually assess the effects of instruction to refine their teaching.		
18. Teachers collaborate to develop/refine the academic curriculum.		
19. Teachers use methods such as cooperative learning, peer tutoring, and computer-assisted instruction to promote learning success for all students.		
20. Teachers participate in professional development activities to keep up-to-date on instructional practices.		
	12. I learn "a lot" in most of my classes.	12. My child is learning "a lot" in most of his/her classes.

High Expectations for Student Achievement	High Expectations for Student Achievement	High Expectations for Student Achievement
21. All students are expected to learn a full range of skills—from basic memorization to complex problem solving.	13. All students are expected to learn a range of skills—from memorization to problem solving.	13. My child is expected to learn a full range of skills—from basic memorization to complex problem solving.
22. Teachers believe that all students can master the basic skills.	14. My teachers believe that all students can master the basic skills.	14. My child's teachers believe that all students can master the basic skills.
23. Teachers clearly inform students and parents of what students are expected to know and be able to do by the end of the unit or semester.	15. My teachers explain what students are expected to learn by the end of the unit or semester.	15. Parents are informed of what their child is expected to learn.
24. School standards are both challenging and attainable.	16. School work is challenging.	16. The school's expectations for student learning are challenging.
25. All staff have high expectations for student achievement.		17. My child's teachers expect students to do well.
26. All staff believe that students can learn regardless of their ability.		18. My child's teachers believe that students can learn regardless of their ability.
27. Teachers assume responsibility for student learning.	17. My teachers try very hard to help all students learn.	
28. Students are encouraged to set high learning goals for themselves.	18. My teachers encourage students to set high learning goals.	
29. Teachers foster the development of independent learning.		
30. Time spent in pull-out programs is expected to be short and effective.		

(continued)

APPENDIX Continued

TEACHER/STAFF SURVEY	*STUDENTS*	*PARENTS*
Frequent Monitoring of Student Progress	**Frequent Monitoring of Student Progress**	**Frequent Monitoring of Student Progress**
31. Teachers often give students feedback on their progress.	19. Teachers often let me know how I am progressing.	19. Teachers often give students feedback on their progress.
32. Teachers promptly evaluate and return homework.	20. Teachers promptly grade and return homework.	20. Teachers promptly grade and return homework.
33. Teachers diagnose academic problems early.	21. My teachers know right away when students have trouble learning.	21. Teachers diagnose academic problems early.
34. Teachers clearly explain their grading systems.	22. My teachers explain their grading systems.	22. Teachers explain how my child's work is graded.
35. Teachers give clear explanations before assigning seatwork or homework.	23. My teachers give clear explanations before assigning seatwork or homework.	23. My child's homework is clearly explained.
36. Clear classroom standards for student behavior are used consistently throughout the year.	24. My teachers have specific classroom rules with clear consequences when rules are broken.	24. Discipline problems are handled quickly with fairness.
37 Students are given an active role in assessing and evaluating their own progress.		
38. Teachers use tests and other forms of assessment to evaluate student learning.		
39. Information from monitoring students' progress is used to adapt instruction to meet individual student needs.		
40. Results from students' progress are used to plan weekly instruction.		

Positive School Climate	**Positive School Climate**	**Positive School Climate**
41. The school is clean and comfortable.	25. The school is clean and comfortable.	25. The school is clean and comfortable.
42. People feel safe at this school.	26. People feel safe at this school.	26. People feel safe at this school.
43. The school staff really cares about students.	27. My teachers really care about students.	27. My child's teachers really care about students.
44. Students in our school want to learn.	28. Students in our school want to learn.	28. Students in our school want to learn.
45. There is an "aloha" spirit with a feeling of "ohana" (family) in this school.	29. There is an "aloha" spirit in this school.	29. There is an "aloha" spirit in this school.
46. Teacher-student interaction is positive.		
47. Teachers enjoy teaching at this school.	30. Students enjoy coming to school.	30. My child enjoys going to school.
48. Discipline problems are handled with fairness, emphasizing behavior, not personality.		
49. Classroom environments stimulate learning without undue pressure.		
50. The school staff works cooperatively together.		
Positive Home-School Relations	**Positive Home-School Relations**	**Positive Home-School Relations**
51. Regular, frequent home-school communications are maintained.	31. The school communicates regularly with Parents.	31. The school communicates regularly with parents.
52. Parents often receive information about students' progress.	32. My parents often get information about my progress in school.	32. Parents often receive information about their children's progress.
53. School events are scheduled to encourage parents' attendance.	33. School events are scheduled to encourage parents' attendance.	33. School events are scheduled to encourage parents' attendance.
54. The staff welcomes parents at this school. (#58)	34. My parents feel welcome at this school.	34. Parents feel welcome at this school.
55. Parents are involved in major decisions about students.	35. The school wants parents to be involved in major decisions about students.	35. Parents are involved in major decisions about students.

(continued)

APPENDIX Continued

TEACHER/STAFF SURVEY	*STUDENTS*	*PARENTS*
56. School staff encourages parents to become involved in activities that support the school's instructional program.	36. The school encourages parents to become involved in school activities.	
57. Parents are offered various options for involvement, e.g., tutoring their children at home, helping in classrooms, joining school councils.		36. Parents are offered various options for involvement like tutoring their children at home, helping in classrooms, joining school councils, etc.
58. The school staff is responsive to parent inquiries.		
59. The school staff continually looks for ways to involve parents, students, and community in decision making.		
60. Teachers let parents know that parent involvement makes a difference in children's school performance.		

NOTES

1. Estimation of random parameters with HLM multilevel regression software (Bryk, Raudenbush, & Congdon, 1996) results in an optimally weighted estimate of the intercept or slope that takes into consideration the data within each unit as well as data from all other similar units. More reliable measures of the random parameter within each unit are afforded more weight in calculating the overall parameter estimate (see Bryk & Raudenbush, 1992, for further discussion of empirical Bayes estimators and residuals). This results in improved estimation across a variety of Level-2 conditions (e.g., unbalanced within-unit sample sizes, small sample sizes). The reliability of the intercept estimates within each unit may be determined from the variance components and the unit sample size (see Heck & Thomas [2000] for further discussion). For the smallest unit (with 6 students), the reliability of the estimates were .54 (reading), .58 (math), and .53 (language). These may be compared with the average intercept reliability coefficients for all units given in Table 3.

2. The state database for the Effective Schools Survey does not include information about return rates for individual schools. Return rates were estimated on a random subset of schools in the study (n = 30) by obtaining the information from the individual schools. The subset of schools ranged in size from 314 students to 1630 students (M = 616). This compared favorably with the overall school mean in the study (M = 605).

3. Researchers have found that two-stage approaches using various estimation methods do not seem to affect the size of the regression coefficients (e.g., Chou, Bentler,& Pentz, 1998; de Leeuw & Kreft, 1986; Tate & Wongbundhit, 1983), although errors in the between-level model may not be as efficient as the one-stage (simultaneous estimation) approach. Chou et al. (1998) compared a single-stage estimation approach using HLM and a two-stage, multilevel structural equation modeling (SEM) approach. They constructed a separate SEM model for each school first (i.e., similar to constructing a Level-1 model for each school in the current study using HLM). The variations in the random parameters at the school level were then used in the school-level SEM. At the second stage, an SEM model was developed with the random parameters as outcome measures and the other school-level variables as predictors. Similar to de Leeuw and Kreft (1986), it was found that the standard errors were larger in the two-stage approach, (which yields fewer findings of significance than the one-stage method), the fixed effects (individual-level predictors) were similar to single-stage estimation, and the significance tests and inferences were quite similar. The major limitation of the approach is that the variability in the Level-1 errors cannot be modeled in the Level-2 model.

For comparative purposes, the data were also modeled using multilevel SEM with Mplus (Muthén & Muthén, 1998) and simultaneous estimation at both levels. There are a few differences worth noting. Multilevel SEM results in group mean-centered (unadjusted) estimates of the within-group model. It is also not possible to model randomly varying slopes across levels simultaneously with existing software (Kaplan, 1998), although the slopes may be partitioned into within- and between-group components. Conceptually, however, the overall pattern of results was very similar to that obtained in the study presented in terms of the statistical significance and impact of the model's parameters at the individual and school levels.

4. The Level-1 model estimated with HLM for each outcome was as follows:

$$Y_{ij} = B_{0j} + B_{1j} \times (\text{Gender}) + B_{2j} \times (\text{Age}) + B_{3j} \times (\text{Lunch}) + B_{4j} \times (\text{Zpreachieve}) + B_{5j} \times (\text{Filipino}) +$$

$$B_{6j} \times (\text{Hawaiian}) + B_{7j} \times (\text{Samoan}) + B_{8j} \times (\text{Sped}) + r_{ij},$$

where lunch = free or reduced-cost lunch; Zpreachieve = previous achievement; sped = special education.

The Level-2 random coefficients are B_{0j} (the adjusted school intercept) and B_{4j} (the slope for previous achievement). The coefficients indicate how the outcomes are distributed in organization *j* as a function of the measured person characteristics. Each random coefficient is conceived as an outcome that is dependent on a set of school-level variables (Bryk & Raudenbush, 1992).

5. In the Bayesian approach, the optimal estimator of a school's intercept or slope is a weighted combination of the within-group and between-group components based on the relative reliability of the group's data. By considering the precision of each group's regression line (see Bryk and Raudenbush [1992] for further discussion of shrinkage), Bayes estimators represent an improvement over the ordinary least squares regression estimates across a variety of conditions found among the schools in the sample (e.g., unbalanced within-unit sample sizes, small group sample sizes). It should be kept in mind, however, that slope coefficients are somewhat more difficult to estimate reliably (Bryk & Raudenbush, 1992) because they generally have greater sampling variability than sample means, which affects their precision of measurement (e.g., they may have larger errors).

6. In the SEM, matrices are used to specify the set of relationships implied in the theoretical model. The measurement model relates the observed indicators to their underlying (latent) factors. The general factor equation (for convenience using a "y" specification in LISREL notation) is

$$\mathbf{y} = \mathbf{7}_{y}\mathbf{0} + \gamma,$$

where **y** is a vector of observed variables (i.e., the six indicators of school quality, the three slope improvement residuals, the three outcome residuals), **7** is a matrix of factor loadings, **0** is a vector of latent factors (i.e., quality indicators, improvement, academic outcomes), and γ is a vector of unique factors (errors). This allows for the incorporation of an individual error term for each observed variable that indicates its reliability in measuring the factor.

The structural model includes the relationships among the latent variables and other predictors in the model. The latent variables to be explained are called endogenous variables (i.e., school quality, improvement, outcomes).They are causally dependent on other endogenous variables or on exogenous variables. Exogenous variables (similar to independent variables) are determined by causes outside of the model and therefore, are not explained by the model (i.e., community socioeconomic status, school size, percentage of sped students). The structural relationships may be written as

$$\mathbf{0} = \quad + \mathbf{\#0} + \ni > + \mathbf{.},$$

where **0** is a vector of endogenous factors, is a vector of intercepts, **#** is a matrix of regression coefficients relating the endogenous factors to other endogenous factors, **∋** is a matrix of regression coefficients relating the exogenous variables (**>**) to the endogenous variables, and **.** is a vector of disturbances (or errors in the equations), indicating that the endogenous variables are not perfectly predicted by the structural equations.

7. The combined model to describe the random intercept is given as

$$Y_{ij} = (\ _{00} + u_{oj} + r_{ij}),$$

which is the same as a one-way ANOVA model with grand mean $_{00}$, a group-level effect u_{oj} and a person effect r_{ij} (Bryk & Raudenbush, 1992). The variability of the outcome is written as

$$\mathrm{Var}(Y_{ij}) = \mathrm{Var}(u_{oj} + r_{ij}) = (\ _{00} + \ ^{2}).$$

The intraclass correlation is then described as the proportion of the variance in the outcome that is between groups:

$$\rho = \tau_{00}/(\tau_{00} + \sigma^2).$$

REFERENCES

Barr, R., & Dreeban, R. (1983). *How schools work*. Chicago: University of Chicago Press.

Bosker, R., Kremers, E., & Lugthart, E. (1990). School and instructional effects on mathematics achievement. *School Effectiveness and School Improvement*, *1*, 213-248.

Bosker, R., & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness research. *International Journal of Educational Research*, *13*, 213-248.

Brookover, W., & Lezotte, L. (1977). *Changes in school characteristics coincident with changes in student achievement*. East Lansing: Michigan State University Press.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Bryk, A., Raudenbush, S., & Congdon, R. (1996). *Hierarchical linear modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software.

Burns, R., & Mason, D. (1998). Class formation and composition in elementary schools. *American Educational Research Journal*, *35*(4), 739-772.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, *8*, 158-233.

Burstein, L. (1992). *The IEA study of mathematics III: Student growth and classroom processes*. Oxford, U.K.: Pergamon.

Chou, C. P., Bentler, P, & Pentz, M. (1998). *A two-stage approach to multilevel structural equation models: Application to longitudinal data*. Unpublished manuscript.

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

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, *64*(1), 5-29.

de Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Education Statistics*, *11*, 57-85.

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

Evans, R. (1999, February 3). The great accountability fallacy. *Education Week*, *28*(21), 35, 52.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Arnold.

Gray, J., & Simes, N. (1991, August). *The stability of school effects over time*. Paper presented at the annual conference of the British Educational Research Association, Nottingham, England.

Gross, S. (1993). Early mathematics performance and achievement: Results of a study within a large suburban school system. *Journal of Negro Education*, *62*(3), 269-287.

Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis*, *17*(3), 323-336.

Hallinger, P., & Heck, R. (1996). Reassessing the principal’s role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, *32*, 5-44.

Hallinger, P., & Heck, R. (1998). Exploring the principal’s contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*, 157-191.

Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education*, *94*(3), 328-355.

Hawai‘i Department of Education (1996). *Effective schools survey*. Honolulu, HI: Department of Education.

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

Heck, R., & Marcoulides, G. (1996). School culture and performance: Testing the invariance of an organizational model. *School Effectiveness and School Improvement*, *7*, 76-95.

Heck, R., & Mayor, R. (1993). School characteristics, school academic indicators, and student outcomes: Implications for policies to improve schools. *Journal of Education Policy*, *8*(2), 143-154.

Heck, R., & Thomas, S. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum.

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

Hox, J. (1995). *Applied multilevel analysis* (2nd ed.). Amsterdam: TT-Publikaties.

Hu, L., & Bentler, P. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation models: Concepts, issues, and applications* (pp. 76-99). Newbury Park, CA: Sage.

Jackson, K. (1982). *Achievement differences in mathematics: A search for causal explanations*. Paper presented at the annual meeting of the National Council of Black Studies, Chicago.

Jerald, C. & Boser, U. (1999). Taking stock. *Education Week*, *28*(17), 81-97.

Jöreskog, K., & Sörbom, D. (1999). *LISREL user’s guide*. Chicago: Scientific Software.

Kahne, J. (1994). Democratic communities, equity, and excellence: A Deweyan reframing of educational policy analysis. *Educational Evaluation and Policy Analysis*, *16*, 233-248.

Kamehameha Schools/Bishop Estate. (1993). *Native Hawaiian educational assessment project: 1993 survey report*. Honolulu, HI: Author.

Kaplan, D. (1998). Methods for multilevel data analysis. In G. A. Marcoulides (Ed.) Modern methods for business research. Mahwah, NJ: Lawrence Erlbaum.

Kaplan, D., & Elliott, P. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling*, *4*(1), 1-23.

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

Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly*, *30*, 498-518.

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

McDonnell, L. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, *17*(3), 305-322.

Monk, D., & Plecki, M. (1999). Generating and managing resources for school improvement. In J. Murphy & K. Seashore-Louis (Eds.), *Handbook of research on educational administration* (2nd ed.). San Francisco: Jossey-Bass, 491-510.

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

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. Wells, England: Open Books.

Muthén, B., Huang, L., Jo, B., Khoo, S., Goff, G., Novak, J., & Shih, J. (1995). Opportunity-to-learn effects on achievement: Analytical aspects. *Educational Evaluation and Policy Analysis*, *17*(3), 371-403.

Muthén, L., & Muthén, B. (1998). *Mplus user’s guide*. Los Angeles: Muthén & Muthén.

National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Washington, DC: Author.

National Center for Educational Statistics (Report No. NCES 98-015). (1997). *Digest of education statistics, 1997*. Washington, DC: U.S. Department of Education.

Nuttall, D., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, *13*(7), 769-776.

Oakes, J. (1989). What educational indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis*, *11*(2), 181-199.

Olson, L. (1999). Moving beyond test scores. *Education Week*, *28*(17), 67-73.

Raizen, S. (1993). Approaches to the science curriculum for grades K-12. In S. Fitzsimmons & L. Kerpelman (Eds.), *Pre-college teacher enhancement in science and mathematics: Status, issues, and problems*. Report submitted to the National Science Foundation. Cambridge, MA: Abt Associates.

Raudenbush, S. (1989). The analysis of longitudinal, multilevel data. *International Journal of Educational Research*, *13*, 721-740.

Raudenbush, S., & Bryk, A. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*, 1-17.

Raudenbush, S., & Sampson, R. (1999). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research*, *28*(2), 123-153.

Reynolds, D., & Cuttance, P. (Eds.). (1992). *School effectiveness: Research, policy, and practice*. London: Cassell.

Reynolds, D., & Packer, A. (1992). School effectiveness and school improvement in the 1990's. In D. Reynolds & P. Cuttance (Eds.), *School effectiveness: Research, policy, and practice*. (pp. 171-188). London: Cassell.

Salganik, L. (1994). Apples and apples: Comparing performance indicators for places with similar demographic characteristics. *Educational Evaluation and Policy Analysis*, *16*(2), 125-142.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results of a reanalysis of the Inner London Education Authority's Junior School Project data. *British Educational Research Journal*, *19*(4), 381-405.

Sammons, P., Nuttall, D., Cuttance, P., & Thomas, S. (1995). Continuity of school effects: A longitudinal analysis of primary and secondary school effects on GCSE performance. *School Effectiveness and School Improvement*, *6*, 285-307.

Scheerens, J. (1993). Basic school effectiveness research: Items for the research agenda. *School Effectiveness and School Improvement*, *4*, 17-36.

Scheerens, J., Vermeulen, C., & Pelgrum, W. (1989). Generalizability of instructional and school effectiveness indicators across nations. *International Journal of Educational Research*, *13*(7), 789-799.

Seltzer, M. (1995). Furthering our understanding of the effects of educational programs via a slopes-as-outcomes framework. *Educational Evaluation and Policy Analysis*, *17*(3), 295-304.

Seltzer, M., Frank, K., & Bryk, A. (1994). The metric matters: The sensitivity of conclusions about growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis*, *16*(1), 41-50.

Tate, R., & Wongbundhit, Y.(1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics*, *8*, 103-120.

Viadero, D. (1999). Setting the bar: How high? *Education Week*, *28*(17), 21-26.

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis*, *20*(3), 137-156.

Wiley, D., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis*, *17*(3), 355-370.

Willett, J., & Sayer, A. (1996). Multilevel models from a multiple group structural equation perspective. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 125-158). Mahwah, NJ: Lawrence Erlbaum.

Willms, J., & Kerckhoff, A. (1995). The challenge of developing new educational indicators. *Educational Evaluation and Policy Analysis*, *17*(1), 113-131.

Willms, J., & Raudenbush, S. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, *26*(3), 209-232.

Witte, J., & Walsh, D. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis*, *12*(2), 188-212.