PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

MSU Is An Affirmative Action/Equal Opportunity Institution

TESTING THE CONNECTICUT EFFECTIVE SCHOOLS CHARACTERISTICS:
AN APPLICATION OF HIERARCHICAL LINEAR MODELING

By

Bruce Alan Brousseau

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling and Educational Psychology

1989

ABSTRACT

TESTING THE CONNECTICUT EFFECTIVE SCHOOLS CHARACTERISTICS:
AN APPLICATION OF HIERARCHICAL LINEAR MODELING

By

Bruce Alan Brousseau

The psychometric properties of the 1984 version of the Connecticut

School Effectiveness Questionnaire (CSEQ) were assessed.

Characteristics of 103 elementary schools in South Carolina were

measured with this instrument to study the relationship between these

school-level characteristics and student achievement as measured by

South Carolina's Basic Skills Assessment Program (BSAP) math and

reading exams at the third grade level.

Hierarchical Linear Modeling (HLM) is used to determine 1) the

school level reliability of the CSEQ scales, 2) what effect these

school factors have on mean student achievement after statistically

controlling for student background characteristics, and 3) how the

measured differences in these factors account for differences in the

relationship between student-level variables and achievement within

South Carolina elementary schools.

Though the CSEQ was found to be a reliable measurement instrument

at the school level for all seven characteristics identified by the

Connecticut Department of Education (1984), further analyses indicated

that the validity of several scales was doubtful. In light of this

fact, only five scales were created from the questionnaire data for use

in subsequent HLM analyses. Four of these scales parallel the original

Connecticut subscales while a fifth combines two of the original scales.

Only one of these five final scales (i.e., PARENT INVOLVEMENT) was found to have a statistically significant relationship with average reading, and none had an effect on average math achievement for the third grade classes in the sample. A relationship was found between the "Clear School Mission" characteristic (labeled INSTRUCTIONAL OBJECTIVES) and our proxy measure of socioeconomic status. As the schools' ratings on INSTRUCTIONAL OBJECTIVES increased, the achievement gap between higher SES and lower SES students (in reading) narrows significantly.

With regard to math achievement we found the "Positive Home School Relation" characteristic had a statistically significant effect on the race slope. But, here the achievement gap between Black students and members of other ethnic groups in math widens as the school's rating on PARENT INVOLVEMENT improves.

These findings lead to the conclusion that either 1) the CSEQ is not sensitive enough to detect important processes that influence achievement levels within elementary schools, 2) the theory upon which a number of school improvement initiatives are based is without grounding, or 3) generalized linear regression models are inappropriate for capturing the structure of school improvement. Any one of the above explanations (or some combination of these) could be responsible for the results of this study. One thing is clear, however, the theory used in this study to define the relationship between these characteristics and achievement is not supported by the evidence provided here.

Dedicated to

P. K.

Jimmy Kijai and Mary Willis who shared their data and advice.

Finally, to my wife P. K. a special thanks. You went beyond love to give me the strength to face the challenges of this task. I only hope I can repay you for the sacrifices you made for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction

In the midst of shrinking fiscal support for public education, this decade has witnessed a ground-swell of school improvement initiatives at the local, state, and national levels. Many of these are based on what has come to be known as "effective schools" research (for reviews see Block, 1983; Edmonds, 1982; and MacKenzie, 1983). In a survey of state programs of school improvement activities Odden & Dougherty (1982) report that a number of states including Alaska, Colorado, Connecticut, Delaware, Maryland, Michigan, Missouri, and Pennsylvania have incorporated results from the effective schools research base into their education improvement policies and programs. These authors also indicate that strategies used in most other states frequently conform to the implications of effective schools research. Miles & Kaufman (1985) report that as of September 1984, 1,750 school districts in 25 states across the nation had established some form of school improvement program, "grounded in a base of research knowledge, mainly about effective schools" (p. 150). And, the use of the effective schools research base to inform policy decisions is still growing.

While many practitioners have come to accept the results of effective schools research as valid, a number of researchers have openly questioned the widespread adoption of the research findings for the purpose of school reform (see, for example, Cuban, 1983; Purkey & Smith, 1983; Rowan, Bossert & Dwyer, 1983; and Zirkel & Greenwood, 1987). These researchers have expressed concern about educators embracing effective schools prescriptions without solid evidence of their value. They argue that fundamental shortcomings in the effective schools research base makes the results ungeneralizable at best and counterproductive at worst.

Some distinctions between research approaches with similar names

In what follows, several names will be used to identify research the present study draws from (e.g., effective schools, school effects, school improvement, and school effectiveness). This is mentioned here to make the reader aware that there are subtle distinctions between these terms though they are often used interchangeably in the literature. There are those who would argue that there really is no distinction (e.g., Burstein, 1987) and in a sense this is true.

What all of these studies have in common is their focus on the relationship between school characteristics and student achievement. No attempt will be made here to compare the merits of these lines of research. However, an attempt will be made to correctly label the sources of information as they emerge in the text that follows. This study can best be understood as a merging of these various lines of research.

Purpose of the Study

The primary purpose of this study is to construct an analytic model consistent with the implicit causal assumptions regarding school effectiveness that has emerged in recent research. This model is based on the pioneering work of the Connecticut Department of Education. Specifically, this study is designed to test assumptions regarding the relationship between several effective school characteristics identified in the Connecticut School Effectiveness Questionnaire (CSEQ) and achievement distributions across three cohorts of third graders in South Carolina's public schools. The "implicit causal assumptions" mentioned above are inferred from descriptions of the Connecticut state wide school improvement project as illustrated in the following passage:

> Even after controlling for student background characteristics and school composition effects, the presence of the seven characteristics will increase student achievement in the basic skills areas. Further, these characteristics are assumed to affect the relationship between SES, race, and gender in such a way as to reduce the achievement gaps across these sub-groups. These relationships are linear and additive (i.e., the more a characteristic is observed within a school the greater will be its effect on student achievement.

Simply put, the problem is to investigate whether statistical relationships exist between several effective schools characteristics, identified in the CSEQ, and student achievement. Demonstrating "cause and effect" relationships, though desirable, is not possible with the available data set. The more realistic goal here is to clearly specify the characteristics that are statistically related to achievement for

third grade students in public schools).

The purpose of this study then is twofold. In the first phase we identify scales that can be used as reliable measures of constructs which are presumed to represent characteristics of effective schools. The second purpose of this study is to estimate statistical relationships between these characteristics and student achievement. The first phase of this study will provide predictor variables that have been identified in the effective schools literature as important for sustaining quality and equity in student academic achievement. In the second phase of the study, the effects of these school level predictors on mean levels of achievement across elementary schools in South Carolina will be assessed. This reflects the "quality" issue mentioned above. The "equity" question involves how the relationship between student achievement and background variables like socioeconomic status and race are conditioned by school characteristics.

By applying an analytic technique that was recently developed to deal with nested or hierarchical data, new insights can be gained as to why student background variables are more influential in some schools than in others while avoiding the ambiguity raised by the use of conventional statistical methods.

Definition of Terms

Hierarchical Linear Modeling

The primary analytic method employed in this study is Hierarchical Linear Modeling (HLM). HLM is a regression technique well suited for

analyzing multi-level data sets. In this study HLM allows us to use regression coefficients estimated within schools as outcome variables in a regression between schools. Thus HLM provides explicit modeling of effects at both the student and school level. In this way all estimated effects are adjusted for individual background and school level influences on achievement simultaneously. For example, student level background characteristics (i.e., race and SES) are regressed on achievement separately for each school in the sample. The regression coefficients that result from this first stage of the analysis are then employed as outcome variables in a regression that uses school composition variables (e.g., percent black, percent gifted, etc.) and several characteristics described below as predictors. This procedure will be explained in greater detail in subsequent chapters.

<u>Definition of Variables</u>

It seems obvious that we are still a long way from a consensus on which characteristics are truly important to school improvement. This is probably as it should be given the complexity of the problems posed. Even within the effective schools research base there are differences, though a core of essential characteristics seem to emerge in a variety of literature reviews. The point here is that different "models" may be appropriate for different schools, and the choice of assessment instruments should reflect that.

Given this warning we should go beyond the general definitions listed below when making inferences from this or any other research. What follows are definitions of the seven effective schools characteristics used by the Connecticut Department of Education (1984)

to guide their development of the CSEQ.

1) <u>Positive school climate (PSC)</u>: There is an orderly, purposeful atmosphere which is free from the threat of physical harm. However, the atmosphere is not oppressive and is conducive to teaching and learning.

2) <u>Clear school mission (CSM)</u>: There is a clearly articulated mission of the school through which the staff shares an understanding of a commitment to instructional goals, priorities, assessment procedures and accountability.

3) <u>Strong instructional leadership (IL)</u>: The principal acts as the instructional leader who effectively communicates the mission of the school to the staff, parents and students, and who understands and applies the characteristics of instructional effectiveness in the management of the instructional program of the school.

4) <u>High expectations (HE)</u>: The school displays a climate of expectation in which the staff believes and demonstrates that students can attain mastery of basic skills and that they (the staff) have the capability to help students achieve such mastery.

5) <u>Opportunity to learn and student time on task (TOT)</u>: Teachers allocate a significant amount of classroom time to instruction in the basic skills areas. For a high percentage of that allocated time, students are engaged in planned learning activities.

6) <u>Frequent monitoring of student progress (FM)</u>: Feedback on student academic progress is frequently obtained. Multiple assessment methods such as teacher-made tests, samples of students' work, mastery skills checklists, criterion-referenced tests and norm-referenced tests are used. The results of testing are used to improve individual student performance and to also improve the instructional program.

7) <u>Positive home-school relations (HSR)</u>: Parents understand and support the basic mission of the school and are made to feel that they have an important role in achieving this mission.

Some of these characteristics parallel the school level predictors in the regression analysis employed in this investigation. The actual

scales used in this study will be discussed at great length in Chapter 3. The reader should compare the definitions associated with these final scales and those provided here by the Connecticut research group.

A second set of school-level predictors, mentioned earlier, are the school composition variables. These simply convey the percentage of students within each school that happen to fall into one of five subgroups. Subgroup membership is a dichotomous assignment (i.e., you either belong to the group or not). The five categories used in this study were race, sex, repeater status (i.e., whether the student was retained in grade), gifted status, and participation in the school free or reduced price lunch program. These variables are then school level aggregates of their student level counterparts.

The outcome measures of interest in this study will be South Carolina's Basic Skills Assessment Program (BSAP) criterion referenced test results for third graders in both reading and math (see Huynh & Casteel, 1983). The measurement properties of all these variables will be described in Chapter 4.

The Problem

Two interdependent problems have already been alluded to. The first involves the difficulty of measuring the "effectiveness" constructs. The second deals with the specification of the relationship between effective schools characteristics and student achievement.

Problem 1. Measurement

Factors associated with sources of the effectiveness constructs and their psychometric properties provide the root of the measurement problem. In their review of school climate assessment instruments Gottfredson, Hybl, Gottfredson & Castaneda (1986) state, "the recent wave of enthusiasm for school improvement created jointly by the effective schools phenomenon and recent calls for school reform has caught educational researchers unprepared to meet the demand for practical assessment tools" (p. 1). One concern raised here is that a large number of assessment instruments are now on the market that provide little or no reliability and validity information. Validity (beyond face validity) in most cases is not studied, or discussed, in relation to these assessment instruments. Reliability information, while more frequently documented, still falls short of standards set for instruments used in educational research and evaluation. Even when reliability and validity data are available, they are usually associated with a narrow range of characteristics (see Gottfredson et al., 1986; Guzzetti, 1983).

The constructs are also difficult to pin down because lists of variables change somewhat from one effective schools study to the next. Instruments based on this research that appear to be measuring the same constructs do not. Though this may seem troublesome, an even more serious source of confusion exists within published lists of characteristics. The reference here is to the fact that the same label is often used to define different constructs. An example of this is drawn from one of the most enduring characteristics to emerge from

effective schools literature (i.e., strong instructional leadership).
CSEQ items associated with the construct focuses entirely on actions of
the school principal. Other studies have shown that the principal need
not be the primary instructional leader in a school. Stringfield &
Teddlie (1987) report that they "visited more than one effective or
improving school in which the actual instructional leader was not the
principal, but a faculty member or informal leadership team" (p. 8).

A related problem can be illustrated again with the leadership
characteristic. This turns on the question of, just what does an
effective leader do? Dozens of studies have been conducted to examine
the principal's role in school improvement, yet there is still little
agreement on the matter (see Murphy, 1988; Zirkel & Greenwood, 1987).

The point here is not to downplay the importance of the principal
in creating and/or maintaining school effectiveness, but to caution
those who might want to study this role and its connection to student
achievement. Similar problems arise with regard to other constructs
reported to be measured by the CSEQ. These will be addressed both in
Chapter 2 where recent work involving the seven characteristics studied
here is discussed, and in Chapter 3 where the actual scales used in
this study are described.

Problem 2. Specification of the Relationship between Effective Schools
Characteristics and Student Achievement

The problems associated with specifying relationships between
effective schools characteristics and achievement have traditionally
come from two areas. One source of this problem involves the
theoretical framework upon which effective schools research is built.

The second part of this problem is methodological. The issue surfaces because of the "nested" nature of the data used in school effectiveness studies. As we shall see, methodological problems are often cast in terms of threats to valid statistical inference. From the presentation of the structural relationship problem thus far, the reader might be lead to believe that the theoretical/conceptual and methodological issues are independent of each other. But, of course, this is not the case.

Theoretical/conceptual framework. For the purposes of the present study, the term "theoretical framework" is used loosely to denote a set of hypotheses about the relationship between the effective school characteristics measured by the CSEQ, and reading and math achievement scores of students in the sample.

The investigation of this relationship is made difficult largely because effective school research lacks an explicit theoretical framework. As Sirois & Villanova (1982) point out:

> The research on the characteristics of Effective Schools has, by design, avoided theoretical foundations, choosing instead various "shotgun" research methods, including ex post facto designs, simple correlational analyses, and the generation of hypotheses in an ex post facto manner (e.g., Edmonds 1981; Lezotte 1980; Denham & Liberman 1980).
> This approach to research and practice, in the absence of any theoretical foundations, is already presenting problems for both the advancement of research on the characteristics of effective schools and the implementation of these characteristics in public school settings. (p. 3)

This sentiment is shared by a number of critics who argue that unimpeachable evidence establishing the relationship between effective schools characteristics and student achievement is practically nonexistent (Purkey & Smith, 1983, 1985; Rowan et al., 1983). This

problem arises as a result of weak theory, but has independent sources as well. For example, many school improvement projects seem to pursue the characteristics being studied here for their own sake and not because of any demonstrated link with increased student achievement. Though this may be appropriate in some situations, research that directly examines the connection between the seven school characteristics described in the CSEQ and student outcomes is still needed. Without such research, policy makers will have to rely on other sources of evidence as the basis for their judgments.

Threats to valid statistical inference. Researchers trying to detect links between school characteristics and student outcomes have also been plagued by methodological concerns involving the analysis of multilevel (nested) data. The issue here involves how we might reliably estimate the effects of school characteristics on educational performance across different levels of analysis. Burstein (1980) says, "Attempts at cross-level inference (e.g., using school-level data to infer about individual behavior) generally cause problems" (p. 161). Many of these problems are due to threats to valid statistical inference introduced by methods of analysis traditionally used in these studies. In the view of Raudenbush & Bryk (in press), threats to valid statistical inference arise because of a "mismatch between the reality we seek to study and the analytic tools we have employed to study that reality" (p. 6).

Raudenbush & Bryk (in press) discuss three threats to valid statistical inference that have traditionally hampered analysis of hierarchical or nested data. These threats can arise as a result of

1) choosing an inappropriate level of analysis, 2) aggregation bias, and/or 3) regression slope heterogeneity. Until very recently researchers had only inadequate means of addressing this situation. Later we will see how HLM can be used to overcome some of the obstacles presented by multilevel data analysis.

Raudenbush & Bryk (in press) refer to our first problem as the threat of misestimated precision. Typically this threat to valid statistical inference arises when investigators attempt to assess the effects of a treatment program which has been implemented in a group setting such as a school. When students are sampled as the unit of analysis (ignoring the influence of the school characteristics), the assumption of independent observations required by the analytic model (e.g., ANOVA) will be violated. In this case the estimated standard error of the treatment effect estimate will be negatively biased (e.g., increasing the chance of labeling a new program effective when it actually is not).

To circumvent this problem, some school effectiveness researchers have used the school as the unit of analysis. The typical reasoning here is that since the "treatment" is implemented at the school level, the outcome should enter the analysis at that level as well. The outcome of interest would be the average achievement for all students in a particular school. This practice is likely to conceal some very important aspects of the data, however. For example, a school may seem to be effective in teaching basic math skills even though a specific subgroup of students fail to demonstrate adequate academic progress in that area.

Another shortcoming of indiscriminately using the school mean as the dependent variable is that it precludes the use of student level characteristics, such as race or socioeconomic status, as covariates in the statistical model. This restriction can drastically diminish the statistical power of the test used in the analysis.

Finally, aggregating the data to the school level effectively reduces the nominal sample size. The obvious consequences of this practice is once again a reduction in the power of the statistical test used. Here we are more likely to commit a Type II error (i.e., not finding an effect that may actually be present).

Robinson (1950) was among the first to document the threat to valid statistical inference arising from aggregation bias. He demonstrated that the correlation between two variables can change substantially depending on the level of aggregation of the data used in the analysis. More recently Cooley, Bond & Mao (1981) have shown that the correlation between school mean student characteristics (e.g., race, sex, prior ability, and socioeconomic status) and school mean achievement for the same sample of students is far larger than the correlation between these same variables at the student level. Raudenbush & Bryk (in press) point out that:

> Such differences between correlations occur because variables can take on different meanings at different levels of aggregation, because variables are measured with different degrees of precision at different levels of aggregation, and because of the non-random process by which students are assigned to schools . . . School mean social class will be more precisely measured than individual student social class if the sample of students per school is large. And because students are not randomly assigned to schools, school mean social class will be correlated with other important school variables which may not be measured, such as academic

expectations of a school and the resources available for
instruction.

. . . the correlation between student social class and
achievement can easily be shown to be a weighted combination
of two other correlations: the within-school correlation and
the correlation between school mean socioeconomic status and
school mean achievement (Knapp, 1977). The weights accorded
each of these depends on how much of the variation in social
class is between schools. Hence, analysis at the student
level cannot liberate the researcher from aggregation bias.
To study the relationship between student socioeconomic status
and student achievement in a meaningful way requires
decomposition of this relationship into its between and
within-school components and specification of the confounding
variables at both levels of aggregation. (pp. 15-16)

Of course, this approach can be extended to study the relationship

between any number of student level variable and achievement as is done

in this study.

The third and final threat to valid statistical inference to be

discussed here involves the fact that the influence of covariates often

differs across schools. When this occurs the data will violate the

assumption of homogeneity of regression. In school effects research it

is this same variation that is often at the center of the

investigation. That is, instead of looking upon regression slope

heterogenity as a nuisance in our analysis, it should be seen as the

very thing we wish to explain. Resolving this methodological problem

is essential to answering the "equity" question. That is, why does the

relationship between student background and achievement changes as a

function of the school attended? The perspective taken here is that

school level characteristics influence not only the mean level of

achievement but the entire distribution of student achievement. As

Raudenbush & Bryk (in press) point out:

The distribution of outcomes as a function of ability, sex, race, and socioeconomic status, for example, is parameterized by the partial regression coefficients of these variables within each school (Burstein, Miller, and Linn, 1981). Hence, an adequate analysis requires that the investigator first estimate how much these coefficients vary across schools and then seek to explain why they vary. (p. 19)

The Hierarchical Linear Modeling (HLM) approach was selected for this study because it offers a solution to several major shortcomings revealed by earlier research. With HLM the "unit of analysis" is no longer an issue (i.e., both the student and school data are modeled simultaneously). Further, HLM provides the means for decomposing the relationship between student background characteristics and achievement into between and within school components, effectively eliminating the problem of aggregation bias. Finally, because HLM decomposes relationships between variables into within and between school components, it helps us estimate the degree to which these covariates vary across schools.

Research Questions

This dissertation explores a major portion of the effective schools research base from a new perspective. The research questions are addressed with a three part strategy of 1) using a well known set of variables representing effective schools characteristics, 2) providing evidence for the reliability of these variables at both the teacher and school levels, and 3) explicitly examining relationships between these variables and student achievement. Part one of this approach was addressed when members of the School Council

Assistance Project at the University of South Carolina chose the CSEQ
for their school assessment instrument. According to Gottfredson
et al. (1986) of those instruments designed for use at the elementary
level that provide quantitative evidence of reliability and validity,
the CSEQ has "the broadest coverage of the effective schools
characteristics" (p. 13). The second and third parts of this research
strategy are cast in terms of the research questions outlined below.

Basic Research Questions Associated with the Measurement Problem

Regardless of the analytic approach used, precise measurement of
key variables is essential to produce valid results in any study. In
this investigation we want to answer two questions that are central to
the measurement issue. They are:

1) Can the characteristics that are reported in the effective
   schools literature to improve student achievement be reliably
   measured with a paper and pencil questionnaire?

2) What evidence do we have for the construct and criterion
   validity of these scales?

The "Quality" Issue

The question of whether our school characteristics have a
significant impact on achievement as discussed earlier will be
addressed by the following research question:

What effect will these characteristics have on mean student
achievement (in both reading and math) after statistically
controlling for student background variables and the effects of
school composition (e.g., percent of the student population on the
free or reduced price lunch program, etc.)?

The "Equity" Issue

The problem of exploring interactions between school characteristics and student background variables is reflected to some extent in the following research question:

How do measured differences in these characteristics account for differences in the relationship between important student background variables and achievement (in both math and reading) within South Carolina elementary schools?

Limitations of the Study

The data used for this study consists of school characteristics measured at a single point in time and student achievement data for three separate groups of third graders. A problem with the data is that the year of measurement of school characteristics does not correspond to the years of measurement for student achievement. The dependent variables, then, consist of student-level achievement measures taken in the spring testing periods of 1984, 1985, and 1986. The school-level variables were measured once between October 1984 and December 1985, but there is no way of knowing the exact year. This has a number of implications for our analysis.

First, the data do not allow us to test the effects of school variables on growth in student achievement since we have achievement data on students at only one point in time. The static or cross-sectional nature of the analysis conducted here can conceal effects the characteristics might have on changes in individual student

achievement. Unfortunately, we can not test this. Moreover, since the school characteristics of interest were measured only once during the time frame of the study, an assumption will have to be made which in essence says these characteristics did not change <u>significantly</u> from one year to the next. This could be a questionable assumption to which available information can only give limited support. Researchers who have discussed educational improvement at the school level (e.g., Dyer, 1970; Fullan & Pomfret, 1977) generally agree that new programs take three to five years to become fully implemented and produce intended effects. The complexity of the innovation and acceptance it receives from school staff are but two of many factors that influence this time estimate. The premise here leads one to conclude that significant improvement would not be expected to result from activities of the School Council Assistance Project staff. It seems reasonable therefore to assume that the argument regarding no significant change in school characteristics applies (at least with regard to intentional program changes). To the extent that this assumption is not met, or to the extent that unknown factors change the level of these characteristics in any given school over time, the independent school level variables in this study will be measured with error. The magnitude and direction of this error will be unknown, but due to the nature of linear regression, the consequences will probably be to make the statistical tests conservative. That is, regression coefficients will be attenuated due to unreliability.

A different linkage problem might also influence the results of this study. Achievement and background data from students could not be

matched with their _teacher's_ responses to the Connecticut School Effectiveness Questionnaire. This situation prevents us from including the classroom as another level of analysis in this study, which in turn means that the model specified may still be inappropriate. The likely impact this will have on the analysis is to make the statistical tests more conservative. That is, to the extent that classrooms vary within schools, the consequences of ignoring classroom effects in the analytic model would be to make an inference about no school effect more likely. The argument here stems from the fact that classroom variation that could be explained in an "extended" model will be subsumed under error variance by the model used in this study. The larger the error variance, the smaller the test statistic (indicating effects) will become, all other factors held constant.[1]

Aside from these limitations to internal validity, questions regarding the generalizability of the results can be raised. Random sampling was _not_ used in this study. Instead the sample consists of those schools which had taken advantage of the services provided by the School Council Assistance Project at the University of South Carolina. As a result one must be on guard against the possibility of selection bias. The issue here is whether the sample for which CSEQ data are

---

[1] This argument could be extended to even smaller units within the school (e.g., reading groups in classrooms, etc.) or to larger units (e.g., the school district). Ideally, data regarding "important" variables should be collected and analyzed at all of these levels to reduce this source of error.

available is representative of a larger population. And further, are the 103 schools used in the study representative of all elementary schools in South Carolina (not to mention in the United States). Comparisons that assess similarities and differences between students in the sample and other third graders in South Carolina is provided in Chapter 4.

Finally, there is some evidence to suggest that effective schools characteristics can have different effects at different grade levels. Firestone & Herriott (1982) argue that differences in basic organizational structures of elementary and secondary schools dictate different approaches to improving school effectiveness. Also the developmental level of the students may interact with school characteristics to produce different results. For example, Iverson, Brownlee & Walberg (1981) found that first graders responded positively to their parents level of interest in school activities, whereas older students are not favorably affected. Clearly, inferences from this study should be restricted to third graders.

Overview

This study is designed to test assumptions regarding the relationship between a subset of effective school characteristics and achievement distributions for third graders in South Carolina's public schools. Principal components analysis was used to reduce CSEQ data collected from South Carolina elementary teachers, and provide guidance in measurement scale construction. Conceptual categories identified by

effective schools research as being important to improving student achievement and to the reduction of achievement gaps among different subgroups of students are then reflected in these scales.

These scales were then subjected to a reliability analysis both at the teacher and school levels. Those scales found to have suitable reliability were used to construct a theoretical school effects model. The framework of this model also included student background variables in a within-schools analysis, and school composition variables in the between-schools regressions. This model is tested with the Hierarchical Linear Modeling (HLM) computer program developed by Bryk, Raudenbush, Seltzer & Congdon (1986).

This program will provide a means of addressing the research questions posed earlier. Further it helps resolve some of the problem associated with analyzing multi-level data by using regression coefficients estimated within schools as outcome variables in a regression between schools. This modified "slopes-as-outcomes" approach will be especially useful for exploring the "equity" issue, which is an important component of school effects research.

CHAPTER TWO: REVIEW OF THE LITERATURE

Historical context

The public's perception of its schools seems to fluctuate with changes in a number of economic and political indicators. In part this occurs because public schools have traditionally been viewed with mixed emotions. To many they are perceived as the only hope for a productive, democratic society, and at the same time are blamed for many of our nation's ills. No doubt, every concerned citizen would want better schools. The question becomes, better in what way, for whom, and at what price?

A good example of the ebb and flow of popular opinion regarding education can be found in the events surrounding the Equality of Educational Opportunity study conducted by James Coleman and his colleagues (Coleman, Campbell, Hobson, McPartland, Mood, Winfield & York, 1966). The Coleman Report, as it is more commonly known, was commissioned by the Civil Rights Act of 1964. Previous studies had been done showing the limited efficacy of schools in overcoming inequality. But, none of these academically oriented studies on social class and schooling received the public attention of the Coleman Report.

During the sixties the political/social climate was such that questions were being raised about the inequality of resources assigned to racially segregated schools in the South. On a broader scale, the

educational system was expected to be an instrument in the "War on Poverty" that would help combat problems of inequality across the nation.

Though the Coleman Report and media coverage of its published findings may have dampened our "faith" in the schools' ability to solve the problems of poverty and inequality, new hope springs up in the form of a reaction to its negative message. Subsequent education movements (e.g., accountability, effective schools, and school finance reform) emerged, in part, to restore this faith.

Contributions of School Effects Research

Prior to the Coleman Report schools were evaluated primarily by the amount of inputs they receive (e.g., per pupil expenditure). The early "input-output" studies provided a basis for Coleman and his colleagues to build on. These early studies demonstrated the importance of including family background (i.e., socioeconomic status) in models of school success. The Coleman et al. (1966) study is thought by many to be the prototype of school effects studies that would follow. It also helped legitimize the use of achievement test scores as a measure of educational output, thus opening the door to educational researchers who felt that this measure was more closely related to the work done in schools.

Both Coleman et al. (1966) and a later reanalysis of that data by Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns & Michelson (1972) found some school effects. In fact the later study reported that about 15% of the total variance in student achievement lies between schools. Some school effects researchers cite this finding as evidence of the

futility of searching for school level processes that affect student achievement, but others have argued that once we more clearly specify our constructs, variables like socioeconomic status (SES) will no longer be the central factor in models of school success.

Reaction to the Coleman Report

The impact of studies like those of Coleman et al. (1966) and Jencks et al. (1972) on the educational research climate is undeniable. Indeed it would seem that the effective schools movement was motivated, in large part, as a reaction to the pessimistic view of schooling suggested by these much publicized "school effects" studies. The findings have often been portrayed in the mass media as demonstrating that schools make little difference in student achievement.

The Coleman Report was thoroughly scrutinized by researchers who found the conclusion that schools "don't make a difference" hard to accept (e.g., see Bowles & Levin, 1968; Mosteller & Moynihan, 1972). These researchers do not refute the general findings that easily measurable differences among schools (i.e., cost per pupil, class size, number of library books, etc.) have little or no consistent relationship to student achievement. However, critics discuss a number of apparent weaknesses in the design of the study. Lezotte, Hathaway, Miller, Passalacqua & Brookover (1980) outline four methodological issues that can "significantly alter the impressions we receive from studying a school or group of schools" (p. 43). These include (a) the existence of contextual effects, (b) the use of socioeconomic status as a proxy for the school learning climate, (c) disagreement over the

proper unit of analysis, and (d) the appropriateness of the measure of achievement. The discrepancy in conclusions from Coleman et al. (1966) and Jencks et al. (1972), to effective schools research, is then explained in terms of these issues. Taking into account these methodological and conceptual considerations, the interpretation of the Coleman Report that schools make little difference in student achievement once socioeconomic status and race are controlled would have to be changed or severely limited.

The fundamental issue of how to measure effects represented, in Miller's (1985) words:

> . . . perhaps the most important long-term finding of the Coleman Report, that is, that resources expended in a school do not have as much influence on achievement as the social-psychological processes and norms that characterize interactions between staff and students within the school. Retrospectively, much of the impetus for the exemplary schools movement can be seen as a search for the social-psychological factors that make a difference in outcomes between schools. (p. 19)

Our "faith" in the schools ability to solve the problems of poverty and inequality, among others, gave this search for "social-psychological factors" top priority on the agendas of many educational researchers. In the eyes of skeptics, some effective schools researchers had been biased by the narrow focus of their pursuit to discover these factors. However, by the time George Weber conducted his investigation to discover characteristics of effective inner-city elementary schools, practitioners (and the general public) were ready to hear the more optimistic message that schools can and do make a difference.

Early Effective Schools Research

According to Good & Brophy (1986), George Weber "conducted one of the earliest studies designed to identify the processes operating in effective inner-city schools" (p. 573). Weber (1971) deliberately selected four "high-achieving" inner city schools serving predominately poor student populations on which to conduct his study. He found that high-achieving schools are characterized by strong administrative leadership, an atmosphere of high expectations, academic time allocations that give priority to basic skills instruction, good discipline, and regular student evaluations. The parallels between these descriptions of effective school characteristics and those tested in the present study are worth noting.

Though the Weber (1971) study has been criticized on several methodological fronts, Sweeney (1982) says his work "provided educators with a point of departure from the devastating Coleman Report" (p. 346). Good & Brophy (1986) point out that, "Despite the limitations of his study Weber was very successful in stimulating others to explore the issue of how schools make a difference in student achievement" (p. 573).

The intention of early pioneers in effective schools research was to show that some schools with high concentrations of poor and minority students can still produce test results at or above the national average. In fact, Edmonds made the argument that if even one such school could be found, the linkage between socioeconomic status and achievement will be broken. Studies with similar designs and purposes were conducted during this early era of effective schools research. As

critics pointed out flaws in the "no control group" designs,

investigators turned to more sophisticated approaches. The strategy

which developed from this exchange was simply to compare "good" and

"bad" schools. The central issue now becomes the identification of

"effectiveness".

A number of techniques have been proposed and used (see for

example Purkey & Smith, 1983; Rowan et al., 1983; and Anderson &

Mandeville, 1986), but none have proven to be universally acceptable.

In fact Rowan (1985) states that, "The best method of measuring school

effectiveness is unknown; therefore assessment should be undertaken in

a spirit of inquiry" (p. 99). Frechling (1982) has demonstrated that

approaches based on absolute gains, trends, and regression methods tend

to produce inconsistent results. Absolute gains and analysis of trends

tend to be biased against schools serving primarily low SES

populations. Some adjustment had to be made for student inputs and

"hard to change" characteristics of the home and school.

In the late sixties Henry Dyer and his associates (Dyer, 1966;

Dyer, Linn & Patton, 1969) first proposed regression analysis be used

as a technique for measuring school effectiveness in response to the

shortcomings detected in other methods. This approach showed promise

because it compensated for the effects of hard to change social

situational variables when distinguishing between more effective and

less effective schools. Dyer, Linn & Patton (1969) have shown that the

result of using different regression methodologies don't always agree.

Nevertheless, the technique was seen as a fair and straightforward

method for developing effectiveness indices utilized by a number of the

more prominent effective schools researchers (e.g., Austin, 1981).
This method relied on regression analysis to identify "outliers" (i.e.,
schools whose students preform much better or worse than would be
predicted given their "hard to change" background characteristics).
But, outlier studies are not without their problems. Questions still
remain with regard to the definition of effectiveness used in these
studies. Also a school that is identified as effective statistically
need not be effective in an absolute sense.

Klitgaard & Hall (1973) built on the work of Dyer as they
conducted the first rigorous large-scale attempt to locate effective
schools. Their strategy was to focus on residuals from the regression
line controlling for only "nonschool background factors that affect
achievement, and assume that the variation remaining after such a fit
represents school effectiveness (and random variation)" (p. 16).
Klitgaard & Hall (1973) explain that "basically, the task is to find
school outliers on achievement scores that are not explained by
nonschool factors or random error" (p. 18). They defined effectiveness
in terms of scores obtained on standardized reading and math tests from
six data sets. One of these was from the Michigan assessment program
1969 to 1971 results for 4th and 7th graders. As Good & Brophy (1986)
point out:

> Although the data support the contention that some
> unusually effective schools exist, the results basically
> support previous research which indicates that the effect of
> schools are small after nonschool factors (SES, aptitude)
> are controlled. The high-achieving schools that were
> identified represented only from 2% to 9% of the sample.
> These schools were clearly unusual and had relatively more
> achievement than schools with comparable populations;
> however, whether they were effective depends on one's
> definition of effectiveness. (p. 572)

Using methods similar to those suggested by Dyer (1966) and outlined by Klitgaard & Hall (1973), the California State Department of Education (1977) designed a study to examine factors that could distinguish between schools where student test scores are unusually high from schools where these scores are unusually low. In contrast to studies which isolate socioeconomic factors, the California study was designed to isolate specific educational factors that were thought to influence achievement. To accomplish this 21 pairs of elementary schools that were essentially the same with regard to predictor variables but very dissimilar in student achievement were selected from approximately 3,500 schools for which the California State Department of Education's computer information system contained sufficient data for the study.

This study illustrates one of the methodological approaches often employed in effective schools research. Many of the major studies that looked for effective schools characteristics included a case study component (see Brookover & Schneider, 1975; Brookover & Lezotte, 1977 for other examples). The difficulties of finding an appropriate statistical model to match the complexities of the typical school setting may have persuaded many to turn to these more ethnographic methods.

In another study that was a direct outgrowth of work done by Dyer and Klitgaard, the Maryland State Department of Education (1978) regressed school subtest scores on the Iowa Tests of Basic Skills against school average Standard Age Scores to identify 18 high and 12 low residual schools. Subsequent case studies of selected high and low

achieving schools lead them to conclude that there were a number of between school differences and "these differences have more to do with the competence of the professional staff of the school than with any other variable identified in this study" (p. 132).

Critiques of the Evidence

While applauding the intentions of effective schools researchers Purkey and Smith (1983) criticized these studies for (a) using narrow and relatively small samples that lack representativeness for intensive study, (b) failing to partial out the effects of background adequately (c) aggregating achievement data at the school level, (d) comparing positive outliers with negative outliers rather than average schools, and (e) using subjective criteria for determining effectiveness. Notice these criticisms focus on the less than ideal manner in which the Dyer method was applied rather than the technique itself, and inherent problems with the case studies approach.

Small sample sizes were a practical necessity for many researchers. As was mentioned earlier, the "identification of outliers" phase of the investigation was often followed by case studies to pinpoint the characteristics that supposedly distinguished between more effective and less effective schools. Questions about the generalizability of results are commonly raised as a criticism of the case study approach. This is not a problem unique to effective schools research.

The second criticism outlined above refers to the fact that not all relevant background variables such as prior ability and socio-economic status are accounted for or measured accurately in many

effective school studies. This criticism focuses on the problem of specification errors that can seriously affect studies that employ regression techniques.[2] Purkey & Smith (1983) correctly point out that if weak or inappropriate measures are used, "differences between high and low outliers will be confounded with student background differences" (pp. 431-432). This can bias the analysis in unknown directions, and is therefore likely to produce misleading results.

Purkey & Smith's third general criticism addresses the problem of aggregation bias. Aggregation bias was discussed in Chapter 1, mostly in terms of its effects on statistical conclusion validity. As was mentioned there, relationships between variables can change considerably when data are aggregated and analyzed at different functional levels. The arguments used for aggregating or not are usually conceptual in nature. A little recognized rationale for this practice seems to rest on practical necessity, however. Only recently have multi-level statistical models become available. Prior to this, researchers generally engaged in single level analyses at either the individual or school level.

The problems with single-level analysis of multi-level data are well known, and Burstein (1980) has demonstrated the benefits of using statistical models that take into account the hierarchical nature of school effects data. When students are nested within schools, Burstein

---

[2] See Berry & Feldman (1985) for a more in-depth discussion of specification errors in multiple regression.

& Miller (1980) provided a paradigm for multilevel analysis. They suggested using within-school regression coefficients as dependent variables in a between-schools model of student achievement. Using this approach, a within-schools regression equation (predicting individual student achievement from student characteristics) is estimated for each school in the analysis. The regression coefficients from this equation are then each modeled separately as outcome variables using school-level characteristics as predictors. Although, this model provides the basic idea behind HLM, the "slopes as outcomes" approach has raises other problems. As Lee (1986) points out, "the major drawback of this interesting concept is that estimation of regression slopes is often done with a great deal of error, particularly if within-school group sample sizes are small" (p. 4). As we will see in the next chapter, HLM's capacity for dealing with this estimation problem through the iterative procedures described by Raudenbush & Bryk (in press), among others, make it the logical choice for the analysis conducted here.

A fourth general criticism of effective schools research design is related to the tendency for effective schools researchers to compare the extreme outliers in their studies (in essence throwing away information from the "average" schools). Rowan et al. (1983) point out several problems with this approach. Basically, when only extreme groups are selected into the research sample the population to which one can generalize the findings is unknown. In this situation tests of statistical significance are invalid. Also, most schools are not extreme, and information that could help improve the majority of them

may be overlooked using this approach. Other generalizability issues will be addressed later.

Finally, Purkey & Smith (1983) observe that "finding a statistically unusual school does not mean that they are unusually effective" (p. 432). Their point is well taken, but seems to focus on the identification of outliers. It is possible that a school with relatively low absolute achievement test scores could be classified as "effective" using the Dyer method. That some schools identified as "effective" by this method produce relatively poor average achievement scores is the critics' evidence that subjective criteria were used for determining effectiveness. This criticism highlights the fact that many of the early effective schools researchers purposely chose relatively poor school districts with high minority concentrations in which to conduct their studies. This biased sampling leads to the situation described above where an "effective school" in a relative sense would be labeled ineffective in some absolute sense.

Researchers have also been criticized for being too subjective during the "case study" phase of many effective schools investigations. Critics might point out that the reason many "different" effective schools studies produce similar results is that their search for characteristics were influenced heavily by previous work. The use of "subjective" criteria for determining effectiveness is a charge that might be leveled at much of what we think of as observational research. This charge could be especially damaging when one understands the need for such investigations to be atheoretical by design. Holding a theory or preconceived notion about what makes a

school effective would no doubt lead one to overlook other (possibly important) school characteristics. The tendency for human beings (including educational researchers) to work toward theory confirmation rather than falsification may also be involved here. The possibility is always present that some effective schools researchers are guilty of having used more subjective criteria than others while conducting their investigations. To the extent that this may have happened, the strength of the corroborating evidence for the characteristics might be in doubt, but this dose not discount the entire research base.

Application of the Research to Practice

By the late 1970's enough evidence had accumulated regarding what came to be known as the "correlates" of effective schools that a number of school districts and State Departments of Education began programs to implement the research findings. Gauthier, Pecheone & Shoemaker (1985) claim that "The Connecticut State Department of Education's School Effectiveness Project, which was launched in 1981, marked the first state wide attempt to improve effectiveness in schools through a systematic change process using valid methods of assessment founded on emerging research and sound practice" (p. 388). Several sections of South Carolina's Education Improvement Act of 1984 endorse the effective schools model as well.

At the national level, The Effective Schools Development in Education Act of 1985 was drawn up. The United States Department of Education began to summarize and disseminate effective schools findings to practitioners. Miles & Kaufman's (1985) directory of programs is a prime example of this effort (also see Consumer Information Center,

1986; U. S. Department of Education, 1986).

All of this occurred in spite of the fact that, as Lezotte (1986) puts it, the usual "rites of passage" were not followed before taking the research to practice. The rites of passage referred to here are experimental studies designed to verify effective schools models. Two reasons can be given for what some might call premature acceptance of the correlates of effective schools. First, many of the recommendations generated by effective schools research fit nicely into the reform and accountability movements of this period. The second reason for quick acceptance seems less political. In addition to offering hope to disgruntled practitioners and school reformers, the recommendations also have an air of "common sense" about them. Whether they were "proven" by carefully conducted research was not as important as how readily they might be accepted by school staff members.

Both proponents of effective schools models and their critics caution practitioners not to view the findings as prescriptions but as guidelines (i.e., a framework for school improvement). Citing the lack of "experimental" evidence, and perhaps being aware of the problems of translating research into practice from their own experience, Purkey & Smith (1983) and Cuban (1983) have warned educators and administrators to exercise caution when implementing effective schools research.

Even leading voices for the movement like Edmonds (1982) recognized the need for future studies to determine whether the correlates of school effectiveness are also the causes. As Marzano, Guzzetti & Hutchins (1984) indicate, a question remains as to whether the variables identified within the many current effective schools

models (including Edmond's) have a causal relationship with achievement.

Implementation Studies

Attempts to use the findings of research in school improvement projects could have provided evidence regarding the validity of the effective schools characteristics. Unfortunately, there seems to be little progress toward combining the efforts of reformers and researchers.

Clark & MacCarthy (1983) report on a school improvement program (SIP) conducted in New York that was the forerunner of the state wide program in Connecticut. These researchers report that:

> SIP schools with greater student achievement have been more actively committed and more loyal to the SIP process. They take their own school improvement plans seriously and have implemented them with much success, particularly the basic skills component. (p. 23)

This gives one the impression that any program will work as long as the school staff support it. Unfortunately, little solid evidence is offered regarding the effect of these programs on student achievement. This is typical of the school improvement studies. The concern is more with "successful" implementation than on evaluating whether the innovation has the desired result. As Good & Brophy (1986) tell us "without descriptions of intended and actual implementation of instructional, organizational, and social processes it is difficult to assess why achievement increases in some schools and not in others" (p. 585).

Project RISE in Milwaukee (McCormack-Larkin & Kritek, 1983) is touted as another example of the successful application of effective

schools research. But, Purkey & Smith (1983) report that only half of the schools associated with this project show increases in achievement. Though this evidence may seem damaging to the case for the value of effective schools characteristics, this need not be so. It could be argued that these studies fail to show a link between school characteristics and student achievement because of implementation failure or distortion not because the innovation itself was faulty. A related problem involves the fact that changes of this nature take time. School improvement involves incorporating changes all at once, and maintaining them over long spans of time. Most researchers have neither the resources or patience to study this process adequately.

Finally, some characteristics may be valued for their own sake. For example, many principals, parents, and teachers may want discipline in the schools simply because a safe and orderly environment makes working there easier (and less dangerous). Though this may be a strong enough reason to support this characteristic it does not provide evidence that achievement scores will be favorably affected. If student achievement is truly what we are concerned with, achievement should be the ultimate criterion of success for these implementation studies.

Recent Work Related to the Seven Characteristics

Work done since the early part of this decade has tried to incorporate and respond to criticisms. The problem of causal ordering

has been addressed by a handful of researchers with little or no apparent success (see Ramey and Hillman, 1983; Marzano et al., 1984; Ramey, 1987). Small sample sizes and high multicollinarity between predictor variables have reduced the usefulness of these reports. Apparently we are still searching for the real causes of student achievement in schools.

Some recent work related to the seven characteristics identified in the Connecticut School Effectiveness Questionnaire (CSEQ) is reviewed below. Though the characteristics are discussed in isolation, we recognize that this does not do justice to the concept of "ethos" currently supported by effective schools proponents. This quality of effective schools is expressed in a recent review which concluded that no single factor accounts for building level success in generating higher levels of student achievement. Instead, school effectiveness research shows that exemplary pupil performance results from a variety of policies, behaviors, and attitudes that together shape the learning environment (Block, 1983).

Safe and Orderly Environment

One would not be shocked to learn that schools which are relatively free from student revolts and open violence fare better on their mean achievement ratings than do schools (with similar student populations) that are in constant turmoil. Weber (1971), Brookover et al. (1979), and Edmonds (1981) all conclude from their investigations that effective schools are characterized by structured learning environments with few disciplinary problems. These authors are quick to point out that "orderly" need not mean "rigid." The labeling of

this construct, and defining its relation to school achievement is still somewhat problematic, however. Designing studies to investigate this linkage is therefore difficult, if not impossible, at the present time.

Lezotte, Hathaway, Miller, Passalacqua & Brookover (1980) acknowledge the confusion and propose the term "school learning climate" be used to refer to this overall construct. They then define school learning climate as "the norms, beliefs and attitudes reflected in institutional patterns and behavioral practices that enhance or impede student achievement" (p. 4). This definition involves a number of other characteristics (e.g., high expectations) discussed later in this section.

On ethical grounds alone it would be difficult to design studies where experimental manipulation of variables that contribute to a safe and orderly environment take place. Thus, there is little experimental support for the claims made for this school characteristic. This is not to say that other forms of evidence do not exist, but the information that we do have does not lend itself easily to an empirical test. Beyond the "common sense" appeal of this characteristic's influence on achievement, several studies comparing private and public schools support the validity of this relationship.

For example, Coleman, Hoffer & Kilgore (1982) found that the "more ordered environment" in private schools was an important factor in explaining their higher achievement. Lee & Bryk (1986) also using High School and Beyond data, reached a similar conclusion. It should be noted that though these conclusions are somewhat tentative. The

fact that they were reached with methods quite different from the research that identified "safe and orderly environment" as an important characteristic for school improvement originally does provide some evidence of the characteristic's validity.

Though there are a number of guidelines telling principals how to create a safe and orderly environment, there has been little or no research to see if changes suggested by these guidelines actually cause changes in student achievement. Researchers may have a problem isolating this construct in that orderliness seems to be related to high expectations, time-on-task, clear school mission, standardization of assessment procedures, and parental involvement in the school.

Clear School Mission

Shoemaker (1984) suggests that Clear School Mission involves both "curriculum alignment" and "instructional alignment". At the heart of this construct is the message that a school's goals, objectives, textbooks, and assessment procedures must be in alignment with each other. Not surprisingly when curriculum is designed and coordinated to more closely match the testing program in a school, test scores will improve. Again Coleman et al. (1982) and Lee & Bryk (1986) provide some evidence that this is true (at least for Catholic high schools). Cohen (1987) concludes, "lack of excellence in American schools is not caused by ineffective teaching, but mostly by misaligning what teachers teach, what they intend to teach, and what they assess as having been taught" (p. 19). But, what evidence do we have for this?

The arguments involve four major elements or linkages. First, norm-referenced achievement tests are not as sensitive to the effects

of classroom instruction as grade level criterion-referenced tests (see

Madaus, Airasian & Kellagahan, 1980; Madaus, Kellaghan, Rakow & King,

1979). As Lezotte et al. (1980) point out "criterion-referenced,

content-specific tests should reflect the greatest influence of

schools, since the objectives on which they are based are specifically

taught by schools" (p. 45). This has been a problem for school

effectiveness research from its inception, and has troubled curriculum

and test designers long before the effective schools movement started.

The second element is that the intended curriculum may not be the

curriculum which is actually taught (Neidermeyer & Yelon, 1981). This

component involves with teacher expectancy effects, poor lesson

planning, unforeseen events in the classroom and school. Floden,

Porter, Schmidt, Freeman & Schwille (1981) conducted a study with 66

fourth grade teachers in which they found that a number of factors

influence the teachers' decisions about what content to actually teach

in their classes. Teachers must consider which textbooks have been

mandated by the district, published objectives, the testing program

used in the school, as well as inputs from parents, principals, and

other teachers. In summarizing this study Porter (1983) states:

> If teachers in a school are to share common goals for
> student achievement, then either these goals must be
> uniformly endorsed by the instructional materials they use
> and the advice they receive from others or there must be one
> source of advice that takes precedence over all others.
> (p. 27)

In either case, establishing common goals will be fraught with

questions of what these goals should be and who (or what) will decide.

The third link addresses the fact that what is presented may not

be what is learned by the students. This point leads to the area of cognitive psychology and refers to learning styles and attending behaviors. Effective schools research does not focus on individual student characteristics in this sense of clear school mission. This linkage is only mentioned here for the sake of completeness.

Finally, the skills being taught may not match the skills being assessed (even on criterion referenced tests). This linkage is central to the research on "transfer" of training. The issue here may be clarified by considering questions like the following:

> Do we test for the same skill when numbers in an addition problem are arranged horizontally as opposed to vertically?

> Should we teach using the exact same format as does the assessment instrument?

Technically speaking, the answer to the first question would be no. If the answer were yes, testing formats would not significantly affect test scores, but they often do. The answer to the second question depends on how important it is to develop the ability within students to transfer what they learn, in specific lessons, to other situations.

Stedman (1985) points out that a number of schools are or have become "effective" only to the extent that they focused on test objectives (sometimes to the exclusion of all else). Porter (1983) says "schools which use tests that are consistent with their instructional focus will appear effective while others will not" (p. 26). Those who use the "narrow focus" criticism seem to be blaming the tests. Regardless of the school's goals, some measurement will be involved to establish that the objectives have been learned to a

satisfactory standard. This implies that both the tests and how they are used should be matters of great concern.

Instructional Leadership

Instructional leadership is seen by many proponents to be the key to school effectiveness. The principal is viewed as the person who plays a major role in establishing and maintaining the other effective schools characteristics. Edmonds (1982) said "there are some bad schools with good principals, but there are no good schools with bad principals" (p. 26).

In a synthesis of eight early effective schools studies, Sweeney (1982) suggests that the research provides evidence that principals of schools with high achievement exhibit particular leadership behavior. However, Glasman (1984) states, "that other studies that investigated both ineffective and effective schools found similar patterns of leadership behavior in both types of schools and even stronger patterns in ineffective schools" (p. 289). A great deal has been written about the leadership role of the school principal (not all of it has been encouraging). Zirkel & Greenwood's (1987) review of effective principals characteristics leaves one with the clear message that at best all of the evidence about the principal's effectiveness isn't in yet. They draw on recent dissertation abstracts to argue that instructional leadership may not have the clear and consistent effect on achievement the earlier research led us to believe.

Bossert, Dwyer, Rowan & Lee (1982) examined the role of the principal as instructional leader. They state that "despite major conceptual and methodological advances in classroom instructional

research during the last few years, little is known about how instructional management at the school level affects children's schooling experience" (p. 34). Difficulties in studying the linkage between student achievement and the principal's leadership characteristics are compounded by the fact that this linkage cannot be a direct one. Glasman (1984) argues that this probably explains the "absence of empirically detected connections between attributes of principals and indicators of student achievement" (p. 289). Murphy (1988) reports that "few of the writings on the principal as instructional leader are reports of research studies" (p. 118). Why then is the idea of leadership effects so pervasive? One reason is that a great deal of the advocacy literature masquerades as research. Another involves the fact that research designs that do focus on the relationship between instructional leadership and student achievement are at best correlational (Glassman, 1984; Ogawa & Hart, 1985). This once again opens the door to the criticism about causal ordering of variables. Bossert et al. (1982) argue "although it is thought that strong instructional leadership facilitates school success, it is equally plausible that the perceptions of strong leadership result from the process of becoming a successful school. The 'black box' and correlational approaches of most of these studies obscures the causes and effects of school structures" (p. 36). In other words, a school's success could easily be attributed to its principal by association.

Finally, the instructional leadership construct may be so difficult to study because of its interdependence with all other factors of school improvement. Bossert et al. (1982) give use a clear

sense of this problem. A careful reading of the items in the Connecticut School Effectiveness Questionnaire that make up this characteristic also reflect the complexity of this construct.

High Expectations

Though school learning climate studies usually contain high expectations as a component, the best support for this characteristic may come from studies of teacher expectations and the self-fulfilling prophecy. Rosenthal & Jacobson (1968) published perhaps the most well known study of the effect of teacher expectations on student test scores. Though this study has drawn fire for its lack of scientific rigor, Rosenthal (1987) reports that "there are over 400 experiments investigating the self-fulfilling nature of interpersonal expectations" (p. 39). Some researchers point out that the original study has never been replicated and use this fact in their arguments to discount the importance if not the very existence of the effect. Others are more enthusiastic about this line of investigation. In a recent review Brophy (1983) says "there seems to be no need for further replications . . . to prove that teacher expectations can have self-fulfilling prophecy effects on student achievement. This has already been demonstrated" (p. 654).

Meyer (1985) counters with arguments similar to those used by critics of effective schools research in general. Three criticisms that stand out are: 1) weak or nonexistent theory, 2) causal ordering problems, and 3) relatively small effect size. Meyer (1985) makes the case that many teacher expectancy studies "are totally lacking any theoretical view and, though possibly helpful for identifying important

variables, these studies do not enhance our understanding of the psychological processes involved" (p. 365).

Mitman & Snow (1985) illustrate the causal ordering problem with a discussion of a study in which Crano & Mellon (1978) employed a cross-lagged correlation approach to establish the direction of the expectancy effect. Using data from over 5,000 elementary school students, Crano and Mellon report that of 84 possible comparisons among variables, 62 were in a direction consistent with the hypothesis that teacher expectancies influence student achievement. However, Rogosa (1980) argues that differences among cross-lagged correlations of the sort studied by Crano & Mellon do not provide as sound a basis for causal inference as previously thought. Rogosa recommended that the cross-lagged correlation approach "be set aside as a dead issue" (p. 257). Mitman & Snow (1985) were only slightly more optimistic that a path analytic approach based on structural regression models could be used to support causal inferences from correlational data of the sort obtained in teacher expectancy research. As these authors point out:

> such developments require strong theory: One must be ready to specify all of the variables entering into the causal paths of interest, and assume that no important variables have been omitted. For teacher expectancy research, this means that variables reflecting teacher and student expectancies will need to be embedded in the much larger network of variables influencing classroom interactions and outcomes. . . It will be some time to come, however, before this research is sufficiently developed to justify strong inference about causal pathways and the role of expectancies in them. (p. 113)

Most reviewers of this research conclude that about five to ten percent of the variance in achievement outcomes can be accounted for in terms of the self-fulfilling prophecy effect. Brophy (1983) indicates

that a 5% effect on educational outcomes is important, but Meyer (1985) seems less enthusiastic on this point. Both authors agree that teacher expectations of student performance are generally very accurate. Thus, Meyer argues, there is little chance for expectancy effects to occur in most cases. That is, low-achievers will typically be recognized as such and treated appropriately.

This line of reasoning brings up a fundamental difference between Meyer's position and that held by effective schools proponents. Whereas the first seems satisfied with the status quo, the latter focuses on changing the situation for low-achievers. The key question is, how should one interact with low achieving students (i.e., students for whom a teacher may hold low expectations)? It is not difficult to imagine a teacher, with the best of intentions, exhibiting behaviors that are debilitating to many students. As Brophy (1983) points out this is more accurately described as poor teaching rather than teacher expectancy effects.

Once again we see the complex and interactive nature of the problem. All of the studies referred to here were of relatively short duration. Research designs were such that the effects had to be mediated by the constant communication of relatively inaccurate expectations, expectations that in naturalistic settings are often contradicted by actual student performance. High expectations may work in "effective schools" because the staff share a general set of beliefs and attitudes that constantly communicate to their students that they can succeed in school. But as Weinberg (1987) points out:

The process by which schools inherit the responsibility

> for social inequity is not well understood. Yet one thing is
> certain - creating high expectations for schoolchildren costs
> less than building new houses and funding new jobs. The
> omnipotence of schooling is a compelling idea in a democracy,
> but sometimes popularity obscures falseness. (p. 35)

The point made here is that people would prefer to think of

expectations as a cheap cure for a deeply rooted and not well

understood systemic problem. This could well be a false hope, but it

may be one that has yet to be given a fair test.

Opportunity to Learn / Time on Task

Since Carroll (1963) published his "model of school learning", a

large number of studies relating time allotments to achievement have

appeared. This work seems to have followed a logical progression as

better methods were developed to study Carroll's model. Early

investigators seemed preoccupied with the relationship between

allocated time and student learning outcomes. As the concept of time

used was refined by research, a more complex association between time

and learning evolved. Rosenshine & Berliner (1978) showed that

allocated time for instruction (i.e., opportunity to learn) was not the

best indicator of effective instruction. There are a number of

influences that weaken the link between time teachers actually intend

to use and time students spend learning.

The phrase "time on task" soon emerged in the literature to

represent a measure of how much time students spent actually engaged in

the study of a particular subject or skill. Fisher et al. (1980) found

that although this measure approximates more closely the actual time a

student spends on some learning activity, it does not provide an

indication of how successful the student may be at learning the task in question. This research encouraged investigators to focus more on the students' ability levels and readiness to learn school subjects rather than just the time dimension involved. Berliner (1979) elaborated this extension of time-on-task by labeling it "academic learning time". By this he meant the amount of time a student spends in a learning activity in which he or she is achieving at a high rate of success. Brophy & Good (1986) summarize research supporting the argument that young students learn best when they cover material at a pace where they make few mistakes.

Berliner & Fisher (1985) claim "that after initial ability is accounted for, no other educational variable is as useful in explaining differences in student achievement" (p. 337). In light of this statement they find it "odd" that time variables are considered trivial by some individuals (e.g., Karweit, 1985 and Rossmiller, 1982). Upon review of the Beginning Teacher Evaluation Study (BTES) and other work relating time and learning Karweit (1983) concludes that "present studies of time and learning, contrary to widely publicized statements, have not produced overwhelming evidence connecting time-on-task to learning" (p. 51). She argues that "these findings point toward an explanation of classroom learning based more on accommodating student diversity in readiness for instruction than on the gross quantity of instruction delivered" (p. 34). This indicates that the confusion expressed by Berliner & Fisher (1985) may be the result of a difference of opinion about what constitutes a <u>time</u> variable.

Frequent Monitoring

Cohen (1982) concludes that "a system for monitoring and assessing pupil performance which is tied to instructional objectives is one of five factors that accounts for differences in effectiveness among schools, when effectiveness is defined in terms of student performance on tests of basic skills" (p. 15). Again there is little evidence to inform us about the mechanics by which frequent monitoring of student performance leads to increased achievement, but two possibilities have been suggested.

One possibility involves the diagnostic function of testing. Mistakes alert the teacher and student to possible breakdowns in understanding. Intuitively, once learning problems are detected they may be addressed with remedial instruction or by adjusting the teaching method used. Note the similarity to Mastery Learning in which a diagnostic test is provided before the mastery test. The items that comprise the frequent monitoring scale of the CSEQ seem to lead one to infer this usage for the school's testing program.

Guided by behaviorist models of learning, some educators have argued that it is important to catch mistakes early so they may be corrected before misconceptions are set in. Frequent monitoring need not mean formal testing, however. Brophy & Good (1986) talk of this construct as useful for "pacing" in many classroom applications.

Others have suggested that the accountability function of testing is more likely to motivate students to do better on achievement tests. Porter (1983) states that within the context of the other characteristics:

the role of testing in creating effective schools seems to fit better with an accountability perspective than with a diagnosis and prescription perspective. Tests are one mechanism available for making clear and forceful the goal of education. (p. 26)

Natriello & Dornbusch (1984) make a similar argument. In their view testing and evaluation sends a clear message to students about what is important in school. This in turn encourages students to increase their efforts in the appropriate direction.

Using what he calls the Individual Learning Expectations (ILE) method with a sample of 387 sixth, seventh, and eighth graders Slavin (1980) "demonstrates that a simple change in the way students are rewarded for their academic performance can significantly increase that performance" (p. 524). The "reward" comes in the form of test scores. He also reports that there was "no tendency for high or low achievers to benefit deferentially from the ILE treatment" (p. 522).

Both the diagnostic and accountability aspects of the frequent monitoring characteristic may be involved to some extent in the effectiveness of schooling. Future researchers should face the task of further refining this construct and isolating its effects on student achievement.

Home School Relations

Though "good" home school relations are implicitly assumed to be beneficial for increasing student achievement, this connection has rarely been the focus of intensive study. Brookover & Lezotte (1977) reported that effective schools usually have more positive parent initiated contacts than do less effective schools. The more qualitative terms "positive" and "parent initiated" are used to modify an

earlier finding that the raw number of parent contacts had no effect on student achievement. Iverson, Brownlee & Walberg (1981) report that parent school contacts improve achievement in young children (i.e., first graders), but seem to have a negative effect on the academic performance of older students (i.e., fourth and eight graders). From their experience Iverson et al. (1981) caution that "increasing the number of contacts without regard to other factors, such as grade, does not produce uniform gains" (p. 395). Obviously this construct has both quantitative and qualitative properties that have yet to be fully accounted for in research.

Due to the nature of the Home School Relations construct, it is not difficult to see that the relationship between it and student achievement may not be linear. There must be a point at which too much contact by parents could disrupt the school's primary function of teaching and learning. Further, the reason for contact between parents and the school may be more important than the frequency. In any case, this construct is still a long way from being fully operationalized.

Summary of Methodological Concerns / Criticisms

A number of methodological weaknesses regarding school effectiveness research have been raised here (see also Purkey & Smith, 1983; Rowan et al., 1983; Cronbach, 1976; Burstein, 1980; and Good & Brophy, 1986). This study addresses only a subset (albeit an important subset) of these issues. What follows is a brief outline of these concerns and how they were address in this study.

Burstein (1980) raised the issue of using inappropriate statistical models to test effects with multilevel data. The nagging problems associated with parameter estimation using nested data were first exposed and studied, with some intensity, by researchers in the school effects field. Recently, several groups of researchers have independently developed software to address these problems. According to Raudenbush & Bryk (in press) "The statistical theories upon which these approaches are based and therefore the statistical properties of their results are, for most practical purposes, identical. All use iterative procedures to compute maximum likelihood estimates of variance and covariance" (p. 24). With these new analytic tools researchers have the opportunity to develop better statistical models for hierarchical data, and thus neutralize concerns over the proper unit of analysis and aggregation bias.

There are other problems identified here that cannot be resolved by the statistical techniques used to analyze the data. Lezotte et al. (1980) criticizes the production function studies for using SES as a proxy for "school learning climate". Purkey & Smith (1983) discuss "failure to partial out background effects" as one problem with the early effective schools research. Both are expressing a concern about model specification, and rightly so. As Achen (1982) says, "without correct specifications, conventional statistical theory gives no assurance that the impact of a variable will be estimated accurately" (p. 11). By directly including measures of the school effectiveness characteristics identified in the CSEQ, along with important background variables into our analysis, part of the concern over specification

errors will be addressed. Specifying the proper functional form of the relationship between these variables and the achievement outcome may not be possible however.

Finally, turning to the outcome measure itself, content validity of tests used in investigations of the linkage between school characteristics and student achievement has become a central issue. Lezotte et al. (1980) and Madaus et al. (1979) both argue that norm referenced standardized tests are relatively insensitive to what is taught in specific courses. For this reason the BSAP tests were chosen as measures of student achievement for this study. The content validity of these tests is assumed to be better than other available achievement measures, because BSAP tests were designed to measure minimal/essential skills, rather than general academic achievement. Therefore, problems associated with poor test alignment should have been minimized.

Summary

The review in this chapter, though by no means exhaustive, does reveal why properly conducted school effectiveness research continues to be so important and challenging. Weak theory, vague and overlapping constructs, and the limitations of working with natural variation combine to provide inconsistent evidence regarding the effects of school characteristics on achievement. This situation would not be as serious as it now is were it not for the fact that so much educational policy is based on this equivocal evidence.

The effective schools characteristics studied here overlap a great deal both in terms of their correlations with each other and in terms of latent causal factors they may represent. Researchers who have ignored this fact probably specify the model needed to explain their results incorrectly. This in turn conceals more about what actually affects achievement in our schools than it reveals. Ethical and practical considerations also weaken research designs, yet even critics agree that the research must continue. Problems uncovered in early studies are being addressed, but the designs and analytic methods used are still in need of improvement.

Bryk et al. (1986) and Raudenbush & Bryk (in press), among others, have provided the tools necessary for more precisely estimating the effects of school characteristics on school outcomes than was possible just five years ago. It appears that in order to do this research "the right way" one must use an analytical model that can accurately assess relationships found in multilevel, multivariate data sets.

Finally, both critics and supporters of the research agree that we are still limited in our knowledge of the actual mechanism involved in making schools more effective. By the time this information filters down to practitioners in the schools this sobering message seems to get lost. Just as summaries of the Coleman Report focused on a bleak picture of public education, more recent summaries of effective schools research tend to downplay cautions made in the original reports. This may raise concerns that the pendulum has now swung too far in the opposite direction. What may be the most important reality for us to

accept when trying to understand this poorly defined collection of simultaneously existing, mutually interacting, ever changing, multilevel variables is that we have only scratched the surface.

CHAPTER THREE: DESIGN OF THE STUDY

Overview

As Chapter 2 discussed, research on school effectiveness has clearly evolved to the point where careful studies are needed to directly examine relationships between the school characteristics identified in recent research on "effective" schools and student achievement. In this chapter, the procedures used in this study to explore these relationships are discussed. The chapter groups the problems of research design encountered in this study into three categories: (1) data collection; (2) measurement; and (3) defining functional relationships among variables.

A brief overview of these topics is in order. Recall from Chapter 1 that this study used data originally collected between October 1984 and November 1985 by the University of South Carolina's School Council Assistance Project, which administered the Connecticut School Effectiveness Questionnaire (CSEQ) to teachers in over 150 elementary schools in the state. These data were then matched to student achievement records for third grade students atttending South Carolina public schools. The student-level data was gathered by the South Carolina State Department of education for the school years 1984

through 1986. As noted in Chapter 1, the CSEQ data were collected only once during the period 1984-1985 while the student achievement data were gathered during the spring testing period of all three years. Thus, in the analyses that follow, data on third graders from the three years under consideration here are pooled and matched to the single year of data for the CSEQ.

The independent and dependent variables for the analysis will be discussed in greater detail below. For now, note that the data set includes measures of individual student achievement in reading and math, measures of various background characteristics of students and measures of various characteristics of schools derived from the CSEQ. Recall from Chapter 1 that a major task of this thesis was the analysis of measurement properties of the CSEQ. This chapter describes how the data from the CSEQ was used to derive a set of scales representing effective schools characteristics.

Once all measures of school and student characteristics were constructed, the analysis turned to an investigation of relationships among variables. In Chapter 2, the need for HLM analysis was discussed, and this chapter describes further the characteristics of this analysis as employed in this study. In particular, the present chapter describes the underlying statistical models that will be used to further analyze the measurement properties of school-level scales derived from the CSEQ, and model for testing the effects of these variables on student achievement in schools.

Data Collection

Archival data were used throughout this investigation. The South Carolina Education Improvement Act of 1984 required that all public schools in the state undergo a comprehensive school needs assessment. The School Council Assistance Project at the University of South Carolina offered such an assessment to a large number of schools. As part of the needs assessment conducted by project staff, CSEQ data were collected and analyzed for later review with cooperating school administrators.

In conjunction with similar legislative mandates South Carolina Department of Education researchers, through their Basic Skills Assessment Program, had been collecting achievement test data from students in South Carolina since 1981. Previous work (e.g., Coleman, et al., 1966; Richards, 1986) had already established the importance of including some measure of socioeconomic status (SES), academic and ethnic background of the student as independent variables in regression equations that model student achievement as the outcome. Anticipating that similar analyses might fall to them, the Department of Education researchers thus collected background information on individual students. Data tapes containing CSEQ, BSAP achievement, and student background information were obtained through negotiations with the School Council Assistance Project director.

The problem was to match the data from these two independent information collection efforts. Data from the 1984, 1985, and 1986 BSAP tapes were matched to data on the CSEQ collected by the School

Council Assistance Project during this period. The final sample consisted of 103 schools and 24,404 students. The data are from the schools that voluntarily participated in the University of South Carolina's School Council Assistance Project. This sample, then, includes only elementary schools having third grade classes in their organizational structures that could be matched with BSAP Department of Education records.

The CSEQ was completed by 2886 staff members of these elementary schools. This sample represents just over one-fifth (22.9%) of all schools serving third graders in the state. To assure anonymity, no students or teachers were identified on the data tapes. Schools were identified by Basic Educational Data System (BEDS) codes only. These numbers were then transformed into arbitrary three digit codes before further analyses are performed.

Measurement

As described in Chapter 2, this study is designed to test the effects of a particular set of school characteristics on achievement. The theoretical interest is in variables measuring school-level characteristics described in recent research on effective schools. These characteristics were measured by data taken from the CSEQ. But, as Chapter 2 discusses, there is a need to control for important student characteristics and for composition variables when analyzing the effects of these effective schools characteristics on achievement. Data on these student characteristics and school composition were taken

from the BSAP data tapes. The preparation of these data for the
statistical analyses undertaken here is described below.

Basic Skills Assessment Program (BSAP) Achievement Tests

Achievement data used as the outcome measure in this study were
collected with South Carolina's Basic Skills Assessment Program (BSAP)
instruments.[3] The BSAP uses criterion referenced achievement test
developed by the South Carolina State Department of Education to assess
educational progress in that state for students in grades 1, 2, 3, 6,
and 8. Scale scores for both reading and math are anchored to the 1984
test.

The objectives in reading for each grade (one, two, three, six,
and eight) are stated in six categories: decoding and word meaning
(DW), main idea (MI), details (DE), analysis of literature (AL),
reference usage (RE), and inference (IN). In math, the objectives are
clustered in five categories: operations (OP), concepts (CN), geometry
(GE), measurement (ME), and problem solving (PS). Six items are
selected for each objective from item banks to create test forms.
Therefore the third grade test in reading is 36 items long, and the
math test contains 30 items.

The development of the reading and math test was contracted with
the Instructional Objectives Exchange (IOX), Los Angeles, California.

---

[3]   Readers interested in a more detailed discussion of the
       development of these tests can contact the Office of Research,
       South Carolina Department of Education for more information (refer
       to the technical report written by Huynh & Casteel, 1983).

Test items were field tested in the spring of 1980, and the first forms were administered statewide in 1981. For each subsequent year, new forms are developed and administered. All test forms have items of similar content, and share a number of common items. This was done so that variations in item characteristics and student ability can be observed from year to year.

A Latent trait model was used to convert the raw test scores to a common scale score system in which the statewide passing criterion is held at 700 for all situations. The standard deviation is set at 100 for both the reading and math test. It is this scale score that is used as the measure of student achievement in the present study.

Background Variables

The BSAP data tapes contained several measures of student characteristics, including sex of the student, race, federal lunch program participation, whether a student repeated a grade, and whether a student was enrolled in a gifted program. These variables were used as indicators of student status and were aggregated to form school-level measures of student body composition.

To help facilitate the interpretation of regression results, all student level predictors were coded as "indicator" variables (i.e., with 0 and 1 coding). After mean scores for the entire sample were determined these individual values were deviated from the sample mean. Now the deviated variables will have a grand mean of zero. This will prove helpful when we attempt to interpret between-school differences in average achievement scores for the typical South Carolina third grader in the HLM analyses. The meaning of the term "typical" as it is

used here will be explained further in Chapter 4.

The final coding of individual background variables is described in Appendix A. Initial coding was as follows: sex of the student (coded 0 for males and 1 for females), race (0 for blacks and 1 for all others), Federal lunch program participation (0 for students on free or reduced price lunch status and 1 otherwise), repeater status (0 for students held back in grade, 1 otherwise), and gifted status (1 if classified as gifted or talented at the start of the school year and 0 if otherwise classified). It should be noted that both repeater and gifted status is determined by local school districts. The decision to so classify students is not based in any way on the outcome measures used here.[4] Also the decision was made to combine racial/ethnic groups in the manner indicated above because preliminary analysis showed that the third largest such group was Asian Americans, with Whites and Blacks accounting for over 99% of the student population.

School composition variables were derived from BSAP records and represent student body characteristics of each school. Examples of these variables are; percent males, percent blacks, percent gifted, percent repeaters, and percent of the students on the free or reduced lunch program (see Appendix A for variable names).

---

[4] This information was obtained from personal communications with Dr. Paul D. Sandifer, Director of the Office of Research, South Carolina State Department of Education.

The Connecticut School Effectiveness Questionnaire (CSEQ)

The ELEMENTARY/MIDDLE SCHOOL Faculty/Staff Attitudes Toward School Questionnaire (CSEQ) was originally developed by the Connecticut Department of Education for conducting needs assessments during that state's school effectiveness project (see Appendix B). This instrument, or some adaptation of it, is probably the most widely used of its kind in this country, yet little is known about its validity or reliability as a measurement tool outside of that project.

The instrument consist of 100 items constructed in such a way that response options are arranged along a five point Likert-type scale (with 1 - Strongly Disagree, 3 - Uncertain / Undecided, and 5 - Strongly Agree). Instructions direct respondents to use the "Undecided" response as "infrequently as possible". Items were not grouped by construct but randomly distributed throughout the instrument with about every sixth item negatively worded.[5]

The Connecticut School Improvement Project staff reported that the scales found in their elementary questionnaire resulted from an extensive review of the available effective schools research (i.e., circa 1979). Table 1 shows the item to scale assignments reported in the Connecticut Department of Education (1984) handbook. Appendix C lists the actual items found in these clusters. Evidence for the

---

[5]   Readers interested in the development of this questionnaire should consult the Connecticut Department of Education (1984) handbook and users' guide. South Carolina used the 1984 version of the CSEQ.

Table 1

Connecticut Item to Construct Specification for the CSEQ.

SCALE	ITEMS IN SCALES
Positive School Climate (PSC)	1 5 8 23 25 44 56 81 87 95
Clear School Mission (CSM)	2 11 15 24 26 29 32 39 41 50 54 59 91 99
Instructional Leadership (IL)	4 6 7 10 13 16 17 19 21 45 46 48 57 60 63 66 70 75 78 82 86 88 90 93 97
High Expectations (HE)	3 12 18 27 38 51 74 77 80 83 85 100
Opportunity to Learn / Time on Task (TOT)	14 28 30 33 52 55 62 65 72 76 89 98
Frequent Monitoring (FM)	20 22 37 40 42 43 53 58 61 73 92 94
Home School Relations (HSR)	9 31 34 35 36 47 49 64 67 68 69 71 79 84 96

Note. Underlined items are negatively worded. See Appexdix C for the
actual items found in these clusters.

empirical soundness of these scales as measures of the constructs found in the Connecticut studies was provided by a Q-sort involving 16 judges who were all familiar with effective schools research evidence. According to the Connecticut Department of Education (1984) handbook, "panel members sorted items into seven categories. In order for an item to remain, it had to be associated with the appropriate character-istic at least eighty percent of the time" (p. 24). This procedure was meant to improve the "construct homogeneity" of the CSEQ scales.

Responses to CSEQ items provided by the South Carolina elementary teachers in our sample were used to recreate the seven scales as defined in Table 1. First, negatively worded items were recoded to match the polarity of other items in the scale. Then the individual scales were formed by adding the scores for all items together with unit weighting and dividing by the number of items in any given scale. This procedure allows us to create scales with a common metric, and retain their interpretability.

As was pointed out in Chapter 2, one type of problem faced by school effectiveness researchers involves the difficulty in defining and isolating variables that represent important characteristics. Efforts to demonstrate causal mechanisms that contribute to school effectiveness have been hampered by high intercorrelations between predictor variables found in the regression models used to explain these relationships. An illustration of this problem was provided by the Pearson product moment correlations among the seven Connecticut scales shown in Table 2.

Table 2

Zero Order Correlations among Connecticut Scales.

	PSC	CSM	IL	HE	TOT	FM	HSR
PSC	-.79[a]	.62	.70	.70	.79	.71	.68
CSM		-.42	.69	.72	.79	.87	.58
IL			-.62	.59	.76	.75	.47
HE				-.27	.71	.70	.74
TOT					-.32	.89	.63
FM						-.34	.60
HSR							-.52

Note.   PSC - Positive School Climate
        CSM - Clear school Mission
         IL - Instructional Leadership
         HE - High Expectations
        TOT - Opportunity to Learn / Time On Task
         FM - Frequent Monitoring
        HSR - Home School Relations

[a]   Numbers in the diagonal represent correlations
      between the mean scale score and its standard
      deviation (n - 103).

Table 2 shows that several rather high zero order correlations were found.[6] These high correlations raise two major concerns. First, the results of this preliminary analysis force us to question whether the Connecticut scales are really measuring independent constructs, and secondly, whether the high collinearity between variables will affect the HLM analyses which will follow. Steps had to be taken to answer these concerns before the HLM models could be tested.

Since we are relying on the perceptions of the South Carolina teachers to provide school-level measures for our characteristics, we decided to subject the teacher-level data to a principal components (PC) analysis. PC analysis was chosen over other data reduction methods (e.g., common factor analysis) because it summarizes data by means of a linear combination of observed responses. The focus here will be on "response homogeneity". As Kim & Mueller (1978) point out, "the mathematical representation of the linear combination of observed data does not require imposing what some may consider a questionable causal model, but it does reveal any underlying causal structure, if such a structure exists" (p. 20). Here the objective is not to explain correlations among variables (as is the case with common factor analysis), but to find the most distinct response patterns (i.e., partitions) in the South Carolina data set.

---

[6] All correlational and principal components analyses were performed using SPSSx. References to this set of statistical packages imply the use of release 3.0 for the IBM VM/CMS operating system.

Since we expected to find seven constructs, we forced a seven component solution on the teacher-level data. Appendix D shows factor loadings after varimax rotation was performed on the seven extracted components. Items associated with each of the seven components are also listed in Appendix D. The scree plot associated with this analysis indicated that five independent dimensions are all that are needed to define the South Carolina data structure.

This analysis indicated that the Connecticut "experts" and South Carolina teachers may not totally agree on the assignment of items to the seven CSEQ scales. Figure 1 gives us some sense of the extent of the mismatch between construct homogeneity and response homogeneity. Notice that the Instructional Leadership items seem to match those loading on the first component (C-1) reasonably well. The same degree of overlap is found for the Clear School Mission scale and component four (C-4). Yet the correspondence between the remaining five scales and components gets progressively worse.

In terms of face validity, a comparison of Appendix C and D reveals that the Connecticut scales make better conceptual sense than the seven components shown in Figure 1. For example, the fifth component (C-5) contains such diverse items as the following:

Mathematics objectives are not coordinated and monitored up through all grades in this school.

Teachers in this school do not hold consistently high expectations of all students.

Criterion-referenced tests are not used to assess basic skills throughout the school.

The common element here seems to be the "negative" wording of scale

Figure 1

Intersection of Connecticut Scales and Seven Principal Components.

PRINCIPAL COMPONENTS

	C-1	C-2	C-3	C-4	C-5	C-6	C-7
PSC	23		5			56 1 8 25	87 81 95 44
CSM		99 91		39 26 59 29 41 54 50 11 2	15 32	24	
IL	17 82 48 7 66 46 60 19 75 78 6 21 90 16 10 88 45 57 97 86 93 13	70			63	4	
HE		77 51	18 74 85 83 80	38 100	27 12		3
TOT		65 76 72 52 28 62			55 98	33 30 14	89
FM	40	53 73 61 43 94 92 22	37		42 58 20		
HSR		34 84 96 71	69 79 67 68 64 36 35 47 9 31		49		

(Left vertical label: CSEQ SCALES)

Note. Entries in cells represent item numbers from the 1984 version of the Connecticut School Effectiveness Questionnaire.

items. The lack of corroboration for the Connecticut scales and questionable construct homogeneity for the five components raised further questions about what constructs were actually being measured with the CSEQ. A second principal components analysis, that restricts the number of orthogonal components to six before rotation, was run to help clarify this situation.

Based on interpretability of the resulting scales, and to a lesser extent the scree plot criteria, we conclude that only five cohesive dimensions exist within the CSEQ data set (see Appendix E). Two of the six components that fell out of this analysis each seem to subsume two of the Connecticut scales. Figure 2 provides an indication of the degree of overlap between the Connecticut item to construct mapping and these six components.[7] For example, component four (labeled C-4 in Figure 2) included a core of items from both the High Expectation (HE) scale and the Connecticut (1984) Home School Relations (HSR) scale. Further, component five from Figure 2 clearly combines items from the Connecticut Time On Task (TOT) and Frequent Monitoring (FM) constructs. Notice that Positive School Climate, Clear School Mission, and Instructional Leadership are matched fairly well with components one, two, and three

---

7   A five component analysis was also run with less satisfactory
    results. That is, the six component analysis was easier to
    interpret.

Figure 2

Intersection of Connecticut Scales and Six Principal Components.

PRINCIPAL COMPONENTS

		C-1	C-2	C-3	C-4	C-5	C-6a
	PSC	1 8 25 44 56 81 87 95		23	5		
	CSM	24	2 11 15 26 29 39 41 50 54 59			91 99	15 32
C S E Q	IL	4		6 7 10 13 16 17 19 21 45 46 48 57 60 66 75 78 82 86 88 90 93 97		70	63
S C A L E S	HE		38 100		3 18 74 80 83 85	51 77	12 27
	TOT	89 30 33				14 28 52 62 65 72 76	55 98
	FM			40	37	22 43 53 61 73 92 94	20 42 58
	HSR				9 31 35 36 47 64 67 68 69 71 79	34 84 96	49

Note. Entries in cells represent item numbers from the Connecticut School Effectiveness Questionnaire. Underlined items were not included in final scales.

[a]The majority of items loading of the sixth component are negatively worded or carry a negative connotation.

respectively. It was decided that the scales from the Connecticut study should be matched with the five scales derived from our second principal components analysis of the South Carolina data. Only items that appear in <u>both</u> scales were used to create the overlapping parallel scales. Component C-6 in Figure 2 was dropped from all subsequent analysis based on lack of face validity. The five remaining scales are a subset of the items found in Table 1, but different labels will be used to distinguish these final scales from the original scales created by the Connecticut researchers. The final scales, and the item numbers associated with each scale, are shown in Table 3.

<u>Final Scale Construction</u>

From this point on the derivation of our final school-level predictors was straightforward. First, responses to negatively worded items were recoded to match the polarity of the other items in their respective scales. The five component-based scales were then formed by adding item scores together (with unit weighting) and dividing by the number of items in the scale. Therefore, the score on any given scale for individual respondents can have a value ranging from 1 (which would indicate that the respondent strongly disagrees with all scale items) to 5 (indicating strong agreement with all items comprising the scale). This approach was taken to maximize both construct homogeneity and response homogeneity of the characteristics being measured while at the same time providing school-level variables appropriate for addressing the research questions posed in this study. A brief description and interpretation of the "final" scales used in this study follows (see Appendix F for items associated with each scale).

Table 3

Item to Construct Specification for Final Scales.

Scale Label	Items in Scale
ORDERLY ENVIRONMENT	1 8 56 81 87 95
INSTRUCTIONAL OBJECTIVES	2 11 26 29 39 41 50 54 59
PRINCIPAL LEADERSHIP	6 7 10 16 17 19 21 45 46 48 57 60 66 75 78 82 86 88 90 93 97
MONITORING & FEEDBACK	22 28 43 52 53 61 62 65 72 73 76 92 94
PARENT INVOLVEMENT	9 31 35 36 47 64 67 68 69 71 79

Note. See Appendix F for actual items associated with each scale.

ORDERLY ENVIRONMENT. The items in the ORDERLY ENVIRONMENT scale generally describe the school's physical condition and atmosphere with regard to student behavior, discipline, and security problems. No reference is made to learning or instruction (see Table F-1). A school with a high rating on this scale can be described as a relatively well maintained, comfortable and safe place to work where students exhibit "positive behavior", follow school rules, and create little or no discipline problems.

Notice that this interpretation differs significantly from the definition of the parallel CSEQ scale "Positive School Climate", in that there is no reference among the items to "purposefulness", teaching and learning, or lack of "oppressiveness". This serves as a further warning not to accept scale "labels" at face value.

INSTRUCTIONAL OBJECTIVES. The nine items that comprise the INSTRUCTIONAL OBJECTIVES scale are a proper subset of the 14 item parallel Connecticut scale labeled "Clear School Mission" (see Table F-2). All nine items deal with sequencing and coordinating the "basic skills" instruction (i.e., reading, math, and language arts). The focus is on curriculum planning and clearly defining instructional objectives in these areas.

Interestingly, only one item in this scale (#29) focuses on mathematics. Four other math items were included in the original Connecticut scale, but did not load heavily on this component in the South Carolina data set. Schools that are rated high on this scale would seem to have well planned and coordinated basic skills curriculum (at least in reading and language arts).

Again differences emerge between this interpretation and the definition provided in the Connecticut Department of Education (1984) handbook. For example, no mention is made of "assessment procedures and accountability" as these terms are generally used. The items do speak to the idea of "clearly articulated . . . instructional goals" for the staff (i.e., teachers), but not of a "mission statement" for students and parents.

PRINCIPAL LEADERSHIP. Twenty of the 21 items that form this scale specifically refer to the "principal" (see Table F-3). Item #86 is the exception that nevertheless could well be referring to another role of the principal (i.e., formal classroom observations). These items describe the frequency of formal classroom observations, follow up to those observations, the expertise of the principal as an instructional leader (or at least resource person), as an interpreter of test results, curriculum coordinator, staff developer, teacher and student motivator, and evaluator.

Ten of these items refer directly to instructional practice while only one (#46) even indirectly mentions managerial skills (i.e., securing resources, arranging opportunities and promoting staff development). Two items refer to lesson plans, and five involve formal classroom observations. On the face of it, this scale might seem to provide a diffuse representation of the principal's role. But, the focus is clearly on the instructional leadership aspect. That is, the higher the rating a school gets from its teachers on this scale the more likely they are to see their principal as a strong instructional leader. The only difference between this interpretation and that of

the Connecticut Department of Education (1984) handbook is that here we
see no direct measure of effectively communicating the mission of the
school to parents or students.

MONITORING & FEEDBACK. This 13 item scale subsumes the
Connecticut Opportunity to Learn / Time on Task and Frequent Monitoring
scale items, and therefore does not represent either construct as an
isolated variable. This point is further supported by the fact that
the correlation between these two Connecticut scales was very strong
(i.e., r = .89 at the school level). The MONITORING & FEEDBACK scale
carries overtones of not only allocating time to important subjects
(i.e., reading and math), but planning instruction in ways that will
eliminate wasted time. The items taken as a whole conveys the message
that widespread use is made of multiple assessment instruments (perhaps
emphasizing criterion-referenced tests), and relatively quick feedback
to students (see Table F-4). A high rating on this scale means that
over 50 minutes and two hours of class time are allocated to math and
reading respectively. Transitions are smooth, success rates kept high,
and work is closely monitored. Feedback is offered regularly both to
students and curriculum planners. One item (#61) speaks to the issue
of using testing to help "plan appropriate instruction". Here again it
is not clear whether testing is used as a diagnostic tool to help
students understand their mistakes or as a motivation device, as was
discussed in Chapter 2.

PARENT INVOLVEMENT. All but two of the 11 items comprising this
scale speak directly of the parents role in monitoring homework,
communicating with teachers, or generally supporting the "mission" of

the school. The other two items refer to homework getting completed on time (see Table F-5). High ratings on the PARENT INVOLVEMENT scale indicate that parents have more contacts with the school, understand and support the schools goals, and encourage their children to do homework. This is similar to the definition given earlier except that no direct mention is made about parents feeling that they have an important role in helping the school achieve its mission.

Overall it seems the "overlapping" items form the Connecticut and South Carolina data, for some scales, can be interpreted in very similar ways to the original CSEQ constructs. To the extent that differences do exist, some of the validity information for the original scales may no longer apply. Nevertheless, the final scales may provide even clearer definitions of the constructs actually being measured. This last point will be especially useful in interpreting the final results of the analyses.

Defining Functional Relationships

The final problem is to specify a statistical model of the relationships among variables. The basic approach used here was Hierarchical Linear Modeling (HLM). Before describing the basic steps, in the HLM analysis, we note a potential problem.

HLM analysis, like other regression techniques, depends on the assumption that variables are normally distributed. However, preliminary analyses indicate that the normality assumption may be violated because both the BSAP achievement tests and the final

component-based scale scores exhibit ceiling effects. As Table 2 shows, the zero order correlations between the final component-based scale scores reflect a moderate to high negative association between their school-level means and standard deviations (i.e., the higher the mean value the smaller the variance). In addition, roughly 80% of the students in the overall sample of 103 schools surpassed the established minimum cut-off score of 700 on the BSAP tests. This means that the distribution on the outcome measure is negatively skewed. The average number of third graders who took the achievement test in each of the schools is 233.2 with a range from 44 to 525 respondents per school. With this sample size, reliance on the Central Limit Theorem to overcome the normality threat seems appropriate.

Rationale for the HLM Approach

Research on school effectiveness has clearly evolved to the point where careful studies are needed to directly explore relationships between the characteristics of schools and student achievement. This requires research designs that are in essence extensions of the school effects studies outlined in Chapters 1 and 2. These early studies focused primarily on relationships between static student background characteristics (like socioeconomic status) and achievement. School level variables used to explain these relationships were typically features like class size, teacher ability, per pupil cost, sector (i.e., private vs. public), and student-teacher ratios. More recent school effects studies have tried to incorporate school process variables and other characteristics derived from policy and practice. These studies are important, but none have directly addressed the

impact of <u>effective schools</u> characteristics, as defined by the CSEQ,

on student achievement or attempted to use these constructs to model

the relationship between student background and achievement.

The subtle "distinctions between research approaches" mentioned in

Chapter 1 may help explain why the effective schools characteristics

have been ignored by school effects researchers for so long.

Statistical techniques similar to the one used here (i.e., HLM) are

gaining favor with the school effects researchers even if the variables

have not. The HLM approach has the potential to provide unbiased

estimates of the effects of these school characteristics on student

achievement after controlling for student background and school

composition effects.

The design of the study is conceptually similar to other

regression analyses, but computationally more complex. Raudenbush &

Bryk (in press) tell us that:

> Since the publication of Burstein's (1980) review,
> several groups of investigators, working independently,
> have developed computational methods and techniques for
> data analysis which essentially resolves the
> long-standing difficulties associated with nested,
> multilevel data. . . these methods share a common core
> consisting of two principles.
> First, they require the investigator to formulate
> explicit structural models for processes occurring within
> each level of a hierarchy. Second, they enable the
> investigator to specify the unique, random effects of
> each unit and to estimate the variances and covariances
> of these random effects. (pp. 19-20)

In this study, we examine processes occurring at two levels, the

within-school level and the between-school level. In the within-school

model, we estimate the effects of student background characteristics on

student achievement. In the between-school model, we estimate the

effect of school characteristics on mean school differences in achievement (i.e., the intercept of the within-school model) and on the regression coefficients representing the effects of the student background characteristics on student-level achievement in each school.

To illustrate this analytic strategy, we draw heavily on the work of Raudenbush & Bryk (in press). The primary purpose of this study is to model Basic Skills Assessment Program (BSAP) achievement test scale scores using both student- and school-level variables as predictors in the regression equations. To simplify the discussion at this juncture, we will refer to a "generic" achievement measure only. The same basic logic applies to both reading and math scale scores.

In the first equation, representing the within-school model, Y_{ij} represents the achievement for student i in school j. Y_{ij} then is seen as a function of measured student-level variables, X_{ijk}, and random error, R_{ij}. Thus the within-school model can be represented by the following equation:

$$Y_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \cdot \cdot \cdot + \beta_{jK-1}X_{ijK-1} + R_{ij} \qquad [1]$$

where the β_{jk} regression coefficients are structural relations that occur within school j which capture the effects of the independent variables on the outcomes. For example, later we will see that the β_{jk} coefficients can represent the distribution of achievement in school j as a function of race or free lunch status.

As Raudenbush & Bryk (in press) point out, "a distinctive feature of HLM is that these structural relationships are presumed to vary

across units" (p. 21). This variation can be estimated by formulating a between-school model. In this model, the outcomes of interest are the k structural parameters, β_{jk}. Raudenbush & Bryk (in press) explain that when k = 0 and the independent variables in equation [1] are centered around their respective group means, β_{j0} represents the mean achievement level in school j. Thus, "the between-school model for β_{j0} affords a direct representation of the effects of school variables on mean achievement" (p. 4). The structural parameters are now seen as a function of school level variables, W_{pj}, and random error, U_{jk} in the following equation:

$$\beta_{jk} = \gamma_{0k} + \gamma_{1k}W_{1j} + \cdot\cdot\cdot + \gamma_{P-1k}W_{P-1j} + U_{jk} \qquad [2]$$

where the Gamma coefficients (i.e., γ_{pk}'s) represent the effects of school level characteristics, W_{pj}, on mean achievement or the structural relations between individual background variables and achievement within-schools. Here the W_{pj}'s represent characteristics such as those measured by the scales derived from the CSEQ or school composition variables. The γ's reflect the effects of these school level characteristics on either mean achievement at a school (β_{j0}) or the race or SES effects on student achievement found in the within-school model.

Statistical Estimation

HLM helps us overcome some drawbacks of the typical slopes-as-outcomes approach when Ordinary Least Squares (OLS) regression methods are used to estimate the outcome variables (i.e.,

the β_{jk}'s). The problem in OLS is that these estimates, $\hat{\beta}_{jk}$, are measured with error. Substituting this estimate into the original equation for β_{jk}, a more complex error structure emerges. Neither the γ coefficients nor the covariance structure among the errors can be appropriately estimated with conventional linear regression methods. HLM provides a maximum likelihood estimation of the covariance structures among this more complex error term, as well as efficient estimates for the γ's. Raudenbush & Bryk (in press) discuss five important properties of parameter estimates generated by HLM.

First, the precision of the $\hat{\beta}_{jk}$ coefficients estimated in any school j depends on the amount of data available on that unit. That is, the contribution of any individual $\hat{\beta}_{jk}$ coefficient is weighted proportional to its precision. This in turn minimizes the effects of sampling variance on inferences about key model parameters.

Second, the estimation procedures are fully multivariate since they take into account the covariation among the β coefficients. Thus, to the extent that these parameters do covary, estimation will be more precise.

Third, the HLM approach allows one to decompose the total observed variance into two parts, the true _parameter variation_, β_{jk}, and _sampling variation_ which arises because $\hat{\beta}_{jk}$ measures β_{jk} with error. Raudenbush & Bryk (1986) inform us that by comparing the estimated parameter variance for each regression coefficient, Var (β_{jk}), to the total variance in the Ordinary Least Squares (OLS) estimates, Var (β_{jk}) + Var ($\hat{\beta}_{jk} \mid \beta_{jk}$), we can derive the reliability coefficients for the random effects in these regression

models. We will see later how this property of HLM can also be used to
estimate the reliability of our outcome measures.

Fourth, HLM allows us to estimate parameter covariance. Estimated
parameter variances and covariances provide the basis for a maximum
likelihood estimate of the correlation between "quality" (the mean
level of achievement) and "equity" (the equality with which achievement
is distributed among various sub-populations).

Finally, HLM draws its strength from the fact that the estimation
of the β_{jk}'s is being repeated across a number of units. The HLM
estimators correct for the unreliability in $\hat{\beta}_{jk}$ as a measure of
β_{jk}. All of these properties speak to the superiority of HLM
relative to more traditional regression methods in applications of the
slopes-as-outcomes approach to studying school effectiveness.

<u>Performing the HLM Analysis</u>

Running HLM is similar in many respects to performing any other
regression analysis. One depends on theory or previous research in
building the regression models. The typical HLM analysis proceeds in
four stages: 1) apportioning parameter variance between and within
schools, 2) assessing the homogeneity of regression assumption,
3) testing for compositional effects, and 4) assessing school effects.
Each stage builds on the previous one. The first of these stages was
used in this study for both achievement outcomes, and to determine the
school-level reliability of the original and final CSEQ scales. Again
the technique will be discussed without regard to particular
achievement measures.

Apportioning Variation

The goal here is to estimate how much of the variation in achievement test scores lies between- and within-schools. The model can be represented by the equation

$$Y_{ij} = \beta_{j0} + R_{ij} \quad \text{where} \quad R_{ij} \sim N(0, \sigma^2) \quad [3]$$

This equation tells us that the achievement test scores for student i in school j are assumed to vary around the school mean, β_{j0}, with variance Sigma squared.

The associated between-school model is:

$$\beta_{j0} = \gamma_0 + U_{j0} \quad \text{where} \quad U_{j0} \sim N(0, \tau) \quad [4]$$

Equation [4] shows that the school means are assumed to be normally distributed around the grand mean, γ_0, of all schools in the sample, with variance Tau.

The HLM analysis provides maximum likelihood estimates of the within- and between-school variance. As was indicated in the previous discussion, with this information we can compute an estimate of the intraclass correlation (ρ_I) using the following formula

$$\rho_I = \tau / [\tau + \sigma^2]$$

This statistic measures the degree of dependence among observations in a school, and it informs us about the estimated proportion of the total

variation in achievement scores which is between schools (with no control for covariates).

This one-way ANOVA model can also provide a statistical test of the between-school variation in school-level predictor variables. HLM provides a Chi-square test of significance to evaluate the hypothesis of no differences in mean achievement or in school-level predictors among schools in the sample.

The analysis also gives us a reliability coefficient (w_j) "analogous to those of classical measurement theory" (Raudenbush & Bryk, 1986). This gives us an indication of the reliability of the school's sample means as estimates of their true means by way of the following equation

$$w_j = \text{Var } (\beta_{jk}) / [\text{Var } (\beta_{jk}) + \text{Var } (\hat{\beta}_{jk} \mid \beta_{jk})] \quad [5]$$

When we add the estimated parameter variance, Var (β_{jk}), for each regression coefficient to its sampling variance, Var $(\hat{\beta}_{jk} \mid \beta_{jk})$ we arrive at an estimate of the total variation in the between-school model. Reliability then can be defined as the ratio of the true parameter variance to the total observed variance. Raudenbush & Bryk (1986) inform us that this coefficient indicates the reliability of $\hat{\beta}_{jk}$ as an indicator of school j's slope. This reliability increases as within-school sample size and between-school variation (i.e., τ) increases.

Assessing Homogeneity of Regression

In the next stage of the HLM analysis student-level variables are added to the within-school model. The goal is to test whether the effects of these student-level variables can be assumed to be constant across schools in the sample. This "unconditional" model provides a baseline for the analyses that follow. In this study the expanded within-school model becomes

$$Y_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \cdots + \beta_{j5}X_{ij5} + R_{ij} \quad [6]$$

where the X_{ij}'s are background variables that were used to classify students according to their sex, race, socioeconomic status, and prior achievement (see Appendix B for coding). The X_{ij}'s are actually the "dummy" values deviated from the grand mean for student i in school j. Thus, β_{j0} is now an adjusted school mean (i.e., the raw school mean minus an adjustment for the mean of the student-level independent variables). The β_{jk}'s each represent the effect of its associated student level variable on achievement test scores in school j.

The between-schools model for the intercept of the within-school model, given in equation [6], is

$$\beta_{j0} = \gamma_{00} + U_{j0} \quad [7]$$

where γ_{00} is the grand mean for the outcome measure. The between-schools model for the other structural relationships in equation [6] are represented by the general formula in equation [8]

$$\beta_{jk} = \gamma_{k0} + U_{jk} \qquad [8]$$

Here γ_{k0} represents the mean within-school regression coefficient for variable k across schools in the sample. The results of this stage of the analysis provide a test of the hypothesis of no variation across schools in their regression coefficients.

$$H_0: \quad \text{Variance } (\beta_{jk}) = 0$$

If this null hypothesis is rejected for any of the k student background variables, that regression coefficient can be said to change significantly as a function of which school the student is enrolled in. Therefore, the coefficient should be modeled as a random effect. If the null hypothesis is retained, but the variable has an effect on the within-school slope, β_{jk}, it can remain in the model as a fixed effect. Here we determine which student-level variables are significantly related to the achievement measure and whether the relationship varies across schools.

Testing for Compositional Effects

School effectiveness researchers often focus on the question of whether a school's composition (in terms of its student body) will have an affect on the achievement of individual students enrolled in that school, beyond the influences already exerted by the student's background variables (Burstein, 1980). To examine this possibility the between-school models can be extended further to include variables that

represent the percentage of students classified into special subgroups
of the school population.

In this study up to five compositional variables are available for
modeling the intercept for the within-school regression equation (i.e.,
mean student achievement). The between-school model for this analysis
is therefore

$$\beta_{j0} = \gamma_{10} + \gamma_{j1}C_{j1} + \cdot \cdot \cdot + \gamma_{j5}C_{j5} + U_{jk} \qquad [9]$$

were γ_{jk} is the effect of school composition variable k on the mean
level of student achievement. C_{jk} is compositional variable k for
school j. The five school composition variables used in this study are
labeled %LUNCH, %MALE, %BLACK, %REPEAT, and %GIFTED.

Assessing Effects of the School Characteristics

The first three stages of the HLM analysis set up the main focus
of this study. By the time we reach the final stage, student
background and school compositional variables which have no effect on
student achievement are dropped from the model. Now the between-school
model is elaborated to include the effects of the school
characteristics measured with the CSEQ. Equation [10] below provides
only an illustrative example of a possible between-schools model for
the intercept from the within-school equation.

$$\beta_{j0} = \gamma_{00} + \gamma_{01}C_{1j} + \gamma_{02}W_{1j} + \cdot \cdot \cdot + \gamma_{06}W_{5j} + U_{0j} \qquad [10]$$

Equation [10], for example, indicates that one compositional and all

five school-level characteristic are being used to model the
intercept. All five characteristic measures will also be used to model
the slopes for race and SES in the within-school model for both math
and reading achievement models. This will allow us to test the
relationship between the five characteristics and average achievement
as well as how these characteristics interact with two important
background variables and their relationship to student achievement.

These equations will take the form

$$\beta_{jk} = \gamma_{0k} + \gamma_{k1}C_{1j} + \gamma_{k2}W_{1j} + \cdot \cdot \cdot + \gamma_{k6}W_{5j} + U_{kj} \quad [11]$$

Where again one composition variable, C_{1j}, and all five effective
schools characteristics, W_{pj}'s, are being used to model the
regression coefficient.

Summary

This chapter described the basic variables and the statistical
model used to estimate relationships among these variables. The final
HLM model contains a within- and between-schools model. The
within-school model predicts student-level achievement from individual
characteristics of students derived from BSAP data. The parameters
from this model then become the dependent variables in the final
between-schools model. The between-schools model estimates the effects
of school composition variables and the five component-based effective
schools characteristics on differences in mean school achievement and

on the estimated effects within schools of individual characteristics on student achievement. As discussed, this analytic approach is superior on several counts to strategies used in past school effectiveness research.

CHAPTER FOUR: ANALYSES AND RESULTS

Outline

Once information on the data tapes have been "cleaned", as discussed in Chapter 3, simple analyses can be performed. These analyses help us understand some basic features of the schools taking part in this study and the student population involved. Recall from Chapter 2 that a question was raised regarding the generalizability of the results of this study to a wider population. The first section of the present chapter will address this concern.

The psychometric properties of the school effectiveness scales used in this study will be explored in the section following the generalizability discussion. Having the variables of interest in place, simple bivariate correlations are then discussed to give the reader a rough sense of the relationship between independent and criterion variables, and the degree of collinearity that may exist in the data set. This chapter concludes with detailed analyses that address the "quality" and "equity" research questions that are the centerpiece of this dissertation.

Preliminary Analyses

Because random sampling was not employed to select the schools that took part in this study the representativeness of the sample used

here might be in doubt. If the students in this sample do not represent some larger population of students, then generalizing the results of this investigation to other educational settings will be problematic. To address this issue comparisons are made between two distinct student groups. The first group represents students in our sample of 103 schools. The second group is comprised of students from those schools not participating in the University of South Carolina's School Council Assistance Project. Comparisons were then made between these two groups on salient background characteristics. For example, Table 4 compares average reading scale scores for third graders in our sample with the remaining South Carolina third graders (i.e., those not in the sample under study) for each year's cohort.

It appears that students in the sample performed somewhat better on the Basic Skills Assessment Program (BSAP) reading test than did other third graders in the state. If we let the students not in the sample represent the population we wish to generalize to (e.g., $\mu - 763.9$ for the 1984 cohort), and compute Z-scores from the basic statistics given in Table 4, for students in each cohort we find $Z_{1984} - 3.59$, $Z_{1985} - 3.56$, and $Z_{1986} - 5.08$. So using a conventional alpha level of .05, at the individual level, a statistically significant difference is observed between the mean Reading Scale Score (RSS) for all three student groups. But, this only serves to illustrate a problem with analyzing data solely at the student level (especially with thousands of cases). A difference of four scale score points can hardly be considered meaningful. When we

Table 4

Reading Scale Scores for Three Cohorts of South Carolina Third Graders.

	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
YEAR	n	Mean	SD	n	Mean	SD
1984	34,235	763.9	90.4	8,657	767.3	88.1
1985	33,315	783.6	91.5	8,939	786.9	87.7
1986	35,439	793.1	91.3	9,492	797.6	86.1

Note. STUDENTS NOT IN SAMPLE entries represent BSAP scores from all third greaders attending South Carolina schools that did not participate in the School Council Assistance Project. All entries are based on student-level data.

observe that the standard deviations for the reading achievement test at this level were all between 86.1 and 88.1, this represents a difference of less than 5% of the standard deviations. Perhaps a more realistic way to think about these comparisons is in terms of meaningful differences.

The same argument can be made in regard to the data on math scores displayed in Table 5. Here a different pattern emerges, however. Notice that, even at the student level, only the scores for the 1986 cohort are statistically different (Z_{1984} = .311, Z_{1985} = 1.41, and Z_{1986} = 3.86). Unlike the reading test scores, which increased steadily from 1984 to 1986, math scale scores fell sharply between 1984 and 1985 then rose slightly again in 1986. Both reading and math achievement in these cohorts differ more from one year to the next than across samples. It would be interesting to learn what factors actually influence these yearly shifts in test scores. Unfortunately, because of the cross-sectional nature of the data we can only speculate about what if any influence our effective schools characteristics might have had on this result.

Table 6 shows the proportion of females in each cohort of third graders. As might be expected, this number is very close to .50 for each year. Also, there is no appreciable difference with regard to sample affiliation. On a school by school basis, the range for the composition variable associated with SEX (i.e., %MALE) was 38.0% to 56.9%. This range is rather restricted in relation to that of the other composition variables used in our analyses.

Table 5

Math Scale Scores for Three Cohorts of South Carolina Third Graders.

	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
YEAR	n	Mean	SD	n	Mean	SD
1984	34,255	799.8	124.3	8,650	800.2	119.6
1985	33,352	778.7	103.2	8,935	780.2	100.5
1986	35,480	782.3	104.4	9,494	786.3	100.9

Note. STUDENTS NOT IN SAMPLE entries represent BSAP scores from all third greaders attending South Carolina schools that did not participate in the School Council Assistance Project. All entries are based on student-level data.

Table 6

Proportion of Females in Populations.

	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
YEAR	n	Mean	SD	n	Mean	SD
1984	34,196	.491	.500	8,650	.484	.500
1985	33,357	.490	.500	8,936	.493	.500
1986	35,465	.497	.500	9,477	.488	.500

Note. Due to the dichotomous nature of the student-level variables, the "Mean" in Tables 6 through 10 is synonymous with the sample proportion.

Table 7 shows that the proportion of Black students in the sample is slightly higher than for other public schools not in our sample, but again the difference is not significant. Also the proportion of Black third graders in our sample does not change significantly from one year to the next. The range for %BLACK across the 103 schools in our sample is much broader than it was for the composition variable %MALE. The minimum value for %BLACK was 0.3% and its maximum value 97.8%.

Table 8 shows the proportion of students who participate in the Federal free or reduced price lunch program in South Carolina schools. This proportion drops slightly from year to year for both populations. The observed trend might help explain the increases in reading test scores noted earlier but does not help us understand the decline in math scores. This provides yet another example of how univariate analyses can be misleading in a number of situations. In terms of student body composition, %BLACK and %LUNCH have very similar distributions across the schools in our sample. Here the percentage of students on the free or reduced price lunch program ranges from 5.9% to 97.8%.

Finally, Tables 9 and 10 give us some insight with regard to the prior achievement of students in our sample. Table 9 compares the proportion of students in our sample who have been held back in the third grade to "repeaters" in other South Carolina schools. Here the proportions are quite different with schools in the sample containing roughly 6% repeaters on average in their student populations as compared to 4.6% for schools not in the sample. It should also be noted that, with respect to school composition, the range for

Table 7

Proportion of Black Third Graders in South Carolina Schools.

YEAR	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
	n	Mean	SD	n	Mean	SD
1984	34,196	.415	.500	8,631	.418	.493
1985	33,326	.414	.500	8,930	.425	.494
1986	35,465	.416	.500	9,488	.421	.494

Table 8

Proportion of South Carolina's Third Graders Not on Lunch Program.

YEAR	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
	n	Mean	SD	n	Mean	SD
1984	34,147	.477	.499	8,639	.489	.500
1985	33,308	.510	.500	8,909	.507	.500
1986	35,441	.518	.500	9,477	.519	.500

Table 9

Proportion of Students Who Repeated the Third Grade.

	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
YEAR	n	Mean	SD	n	Mean	SD
1984	34,000	.045	.207	8,628	.060	.238
1985	33,222	.050	.218	8,878	.067	.250
1986	35,016	.045	.208	9,352	.053	.223

Table 10

Proportion of Third Graders Classified as Gifted or Talented.

	STUDENTS NOT IN SAMPLE			STUDENTS IN SAMPLE		
YEAR	n	Mean	SD	n	Mean	SD
1984	29,795	.039	.194	5,778	.042	.201
1985	29,843	.036	.186	6,004	.074	.262
1986	30,358	.049	.216	6,010	.082	.275

%REPEAT is once again restricted, going from a low of 0.2% to a high value of 21.8%.

Interestingly, the same pattern is observed at the other end of the achievement spectrum. Table 10 shows that there are more students classified as gifted or talented in the sample than in other South Carolina schools.[8] This is especially true for the last two years of data collection. Here schools in the sample have almost twice as many gifted/talented students in their third grade classes as South Carolina schools not in the sample. The range for the associated composition variable %GIFTED is zero to 33.3%. Combining the information provided by Tables 9 and 10 we reach the conclusion that the prior achievement curve for students in our sample is somewhat flatter than that for other third graders in the state. That is, there are more students in the extreme tails of the distribution and correspondingly fewer in the middle.

The information in tables 4 through 10 allow us to state that there is no evidence to suggest that inferential errors will result from generalizing our findings to all third grades in the state of South Carolina. Restrictions on the range of some key school composition variables (i.e., %MALE, %REPEAT, and %GIFTED) may reduce their effects in later HLM analyses, however.

---

[8] It should be noted that Gifted and Repeater status are both determined to an unknown extent by school policy. Therefore, these variables may be poor proxy measures of prior achievement.

Psychometric Properties of Scales

One of the primary aims of this study is to answer questions about
the usefulness of paper and pencil instruments employed in this type of
research. We turn now to an examination of the Connecticut School
Effectiveness Questionnaire (CSEQ) with these questions in mind.

The Original Connecticut Scales

The CSEQ was not originally designed as a research instrument.
Therefore, it seems appropriate to ask if its scales meet acceptable
measurement standards required in the present study (and similar
educational research). We have already seen that the seven scales
created by the Connecticut Department of Education researchers are
highly correlated, suggesting that they do not measure different
constructs, from the teachers' perspective. In this section, we turn
to an analysis of the reliability of these seven scales and then turn
to a similar analysis for the revised scales derived through principal
components analysis of the South Carolina teacher data.

The Connecticut Department of Education (1984) handbook for the
elementary questionnaire offers some evidence for the reliability of
the CSEQ in terms of both internal consistency and test-retest
reliability. Table 11 provides a display of this evidence along with
some additional information derived from analyses of the seven scales
using the South Carolina data set. As this table shows, the internal
consistency estimates (Cronback's Alpha) for the original seven
Connecticut scales range from a low value of .57 for the High
Expectation (HE) scale to a high of .93 for the Instructional

Table 11

Reliability Estimates for Original Connecticut Scales.

Subscale	# of Items	Alpha	Test-Retest Reliability	ρ_I	w_j
PSC	10	.80 (.87)	(.85)	.41	.94
CSM	14	.86 (.90)	(.90)	.28	.90
IL	25	.93 (.93)	(.83)	.37	.93
HE	12	.57 (.55)	(.69)	.23	.88
TOT	12	.67 (.66)	(.74)	.22	.87
FM	12	.78 (.77)	(.67)	.26	.89
HSR	15	.86 (.89)	(.82)	.41	.94

Note. Reliability estimates in parentheses were obtained from the Connecticut Department of Education (1984) handbook. Test-retest reliability was established through administration of the CSEQ to 60 faculty members of one elementary school on two separate occasions ten days apart. 423 teachers provided data for the Connecticut internal consistency estimates. Cronbach's Alpha for South Carolina data was provided by SPSSx.

ρ_I – Intraclass correlations.

w_j – school level reliability estimates from HLM analysis.

PSC – Positive School Climate
CSM – Clear School Mission
IL – Instructional Leadership
HE – High Expectations
TOT – Opportunity to learn / Time On Task
FM – Frequent Monitoring
HSR – Home School Relations

Leadership (IL) scale. This result closely parallels that given in the Connecticut Department of Education (1984) handbook where the Alpha coefficients for these scales are .55 and .93 respectively. Recall from our discussion in Chapter 1 that one of the shortcomings of the Connecticut analysis was that the reliability measures were at the teacher rather than school level. Therefore, one-way ANOVA's were run using the HLM program. These analyses allow us to decompose variance in the data into within- and between-school components and to obtain reliability estimate of our estimates of school means on the seven original scales.

Intraclass correlations were also determined with the aid of HLM as indicators of how much of the total variance for these scale scores lay within and between schools. The column labeled "ρ_I" in Table 11 represents intraclass correlations for these scales. Recall from Chapter 3 that this statistic represents the proportion of total variance in this characteristic that is between schools. This proportion ranges from a low of 22% to a high of 41% of the variance between schools. The final column in Table 11 (labeled w_j) gives the school level estimates of reliability for each scale. This coefficient represents the reliability of the schools' sample means as estimates of their true means. These point estimates of reliability range from a low of .87 to a high of .94, indicating that, at the school level, all Connecticut scales are sufficiently reliable for the purpose of this study.

Scales Derived through Principal Components Analysis

Preliminary principal components analysis of South Carolina data indicated that only five coherent dimensions are identifiable. The items associated with the scales (and their corresponding factor loadings) can be found in Appendix E. These scales produced coefficient Alphas of .82 to .93 during initial reliability analysis (see Table 12). As was reported in Chapter 3, the Time On Task (TOT) and Frequent Monitoring (FM) items seem to cluster under the C-5 group (see Figure 2 for item numbers). The scale labeled C-4 consists of items that were central to the original High Expectations (HE) and Home School Relations (HSR) scales.

In view of the discrepancies mentioned above, it was decided that the final set of scales used to represent the school characteristics under investigation in this study should be formed from the intersection of the original Connecticut scales and the item groupings derived through the principal components analysis of the South Carolina data set. That is, scales from the Connecticut Department of Education (1984) specifications were matched to their closest counterpart derived through the principal components analysis. For example, the scale labeled C-1 in Table 12 was matched with Positive School Climate because this is where the greatest item overlap occurred (see Figure 2). The resulting scale was then labeled ORDERLY ENVIRONMENT. As noted earlier, the fifth group (i.e., C-5) was matched with two different Connecticut scales. Therefore, MONITORING & FEEDBACK (the resulting scale) does not represent either construct independently. The fourth group (C-4) also seems to include a core of High Expectation

Table 12

Reliability Estimates for Principal Components Analysis Scales.

Component	CSEQ Parallel Scale	# of Items	Alpha
C-1	Positive School Climate	13	.82
C-2	Clear School Mission	12	.84
C-3	Instructional Leadership	24	.93
C-4	Home School Relationship	19	.85
C-5		22	.86

Note. Connecticut scales representing Time On Task and Frequent Monitoring were subsumed under the fifth component (C-5). The High Expectation construct had no clear parallel to any scales based on the principal components analysis of South Carolina data.

items, but in this case the match was made with only the Home School Relations scale to maintain its integrity.

This procedure reduces the length of the scales somewhat in every instance. Table 13 gives the reliability information for this final set of scales. As can be seen, the reliability estimates for these shorter scales are comparable to the original Connecticut measures of the effective schools characteristics.

The evidence thus far indicates that the five scales formed from overlapping items retain the favorable properties of the longer Connecticut scales. The appropriateness of the final set of scales defined by these items as measures of the school characteristics central to this investigation is given the added support of confirmation through principal components analysis. With these results for the South Carolina data set, it was decided that the overlapping items would be used to represent the school characteristics of interest in all subsequent analyses. Items comprising these five final component based scales are presented by characteristic cluster in Appendix F. Before going on to answer the remaining research questions, within- and between-schools data sets are created and further simple analyses are undertaken to help us more fully understand the relationships between the variables involved in this study. A further indicator for the similarity of the final scales and original Connecticut scales can be found by examining the diagonal entries in Table 14. Notice the Pearson product moment correlations between the four strictly parallel scales (i.e., ORDERLY ENVIRONMENT, INSTRUCTIONAL OBJECTIVES, PRINCIPAL LEADERSHIP, and PARENT INVOLVEMENT) are all above

Table 13

Reliability Estimates for Final Set of Scales.

Scale	# of Items	Alpha	ρ_I	w_j
ORDERLY ENVIRONMENT	7	.76	.42	.95
INSTRUCTIONAL OBJECTIVES	9	.85	.26	.89
PRINCIPAL LEADERSHIP	19	.93	.37	.93
MONITORING & FEEDBACK	13	.85	.25	.89
PARENT INVOLVEMENT	11	.85	.40	.94

Note. The scale labeled MONITORING & FEEDBACK combines items from the "Time On Task" and "Frequent Monitoring" Connecticut scales. The remaining scales are formed from the intersection of items from parallel component based and Connecticut scales (see Figure 2).

Table 14

Correlations Between Final and Connecticut Scales.

		CSEQ Scales		
Final Scales	PSC	CSM	IL	HSR
ORDERLY ENVIRONMENT	.95	.34	.57	.48
INSTRUCTIONAL OBJECTIVES		.97	.54	.43
PRINCIPAL LEADERSHIP			.99	.43
PARENT INVOLVEMENT				.98
MONITORING & FEEDBACK				

Note. Coefficients in the diagonal in this table represent correlations between Connecticut and final scales as follows: "Positive School Climate" (PSC) with ORDERLY ENVIRONMENT, "Clear School Mission" (CSM) with INSTRUCTIONAL OBJECTIVES, "Strong Instructional Leadership" (IL) with PRINCIPAL LEADERSHIP, and "Home School Relations" (HSR) with PARENT INVOLVEMENT.

The MONITORING & FEEDBACK scale has no Connecticut counterpart.

.95. Secure in the knowledge that our scales represent the effective
schools constructs of interest, we can now go on to explore
relationships between them and other variables used in this study.

Zero Order Correlations

Before we discuss the HLM analysis it is important to get a
feeling for some simple bivariate relationships found in the data set.
This examination of the data with correlational analysis is primarily
meant to help us understand the HLM results.

Correlations at the Student Level

The individual- or student-level background variables constitute
an important class of predictors in school effectiveness research. For
this reason Table 15 was constructed to demonstrate the correlations
between the five background variables available for this analysis, as
well as the reading and math outcome measure (RSS and MSS
respectively).

Some of the results presented here are worth further comment. For
example, notice the relatively high correlation between our proxy
measure of socioeconomic status (LUNCH) and the RACE variable
(r - .54). This indicates that Black third graders in South Carolina
are more likely to be on the free or reduced price lunch program than
are members of other ethnic groups. Also notice that the highest
correlation (r - .57) in Table 15 is between the two outcome measures.
That is, a student who tends to achieve high scores on the BSAP reading

Table 15

<u>Bivariate Correlations for Student Level Variables</u>.

	LUNCH	SEX	RACE	REPEAT	MSS	RSS
YEAR	.03**	.00	.01	.02**	-.05**	.15**
LUNCH		-.01*	.54**	.12**	.27**	.33**
SEX			-.01*	.06**	.00	.07**
RACE				.10**	.27**	.30**
REPEAT					.08**	.12**
MSS						.57**

<u>Note</u>. (n - 24,404), MSS - Math Scale Scores, RSS - Reading Scale Scores.

* One-tailed significance p < .05
** One-tailed significance p < .01

exam can be expected to perform relatively well on the math test.

It is interesting that the LUNCH and RACE variables seem to be better predictors of the outcomes than is the proxy measure for prior achievement (REPEAT). This probably means that REPEAT is an inadequate measure of prior achievement. This conclusion is derived from the fact that prior achievement is usually a better predictor of current achievement than is either SES or ethnic background. We have already noted that REPEAT is not a "pure" measure of prior achievement. It is also possible that the strength of the relationship between REPEAT and the outcome measures is being attenuated by some as yet unknown factor(s).

Finally, it should be pointed out that the correlation between YEAR and the outcome measures corroborate the findings presented in Tables 4 and 5. Recall, YEAR simply indicates the year in which data were collected from each third grader in the sample (i.e., 1984 = -1, 1985 = 0, and 1986 = +1). One would expect therefore that if achievement test scores are increasing over this period (as they are in reading) we should observe a positive correlation (as we do, r = .15) between YEAR and RSS. The opposite trend is revealed for the math data. Thus a small but statistically significant negative correlation is found between YEAR and MSS (r = -.05).

Correlations between the Five School Characteristics and Student Composition Variables

A further illustration of the complexity of this investigation is provided in Table 16. Here the means of our school-level

Table 16

Bivariate Correlations between Final Scales and School Composition
Variables.

	%LUNCH	%BLACK	%MALE	%REPEAT	%GIFTED	SIZE	MYEAR
ORDERLY ENVIRONMENT	-.30**	-.27**	.04	-.21*	.25**	.04	.13
INSTRUCTIONAL OBJECTIVES	-.18*	-.05	.08	.02	.15	-.01	.07
PRINCIPAL LEADERSHIP	-.04	-.04	.12	-.07	.18*	-.01	.16
MONITORING FEEDBACK	-.30**	-.22*	-.11	-.12	.21*	.02	.11
PARENT INVOLVEMENT	-.70**	-.51**	.04	-.35**	.34**	.30**	.20*

Note. All correlations are at the school level ($\underline{n}$ - 103). See
        Appendix A for more detailed descriptions of variables.

*.01 < $\underline{p}$ < .05

**One-tailed significance $\underline{p}$ < .01

characteristics are correlated with the school level values of the composition variables discussed previously. The variable MYEAR, though not strictly speaking a school composition variable, was included in this table for the sake of completeness. MYEAR is the mean value for the variable YEAR in each school. It therefore provides an indication of the proportion of students in each school who were third graders in 1984, 1985, or 1986.

Because the five school effectiveness characteristics were measured only once during the period from 1984 to 1986, one would expect to find no relationship between these variables and MYEAR. However, there is a statistically significant correlation between MYEAR and the scale labeled PARENT INVOLVEMENT in this analysis. This could result, for example, if school administrators knew in some way that they needed help with this characteristic and volunteered early for School Council Assistance Project support. Unfortunately, there is no way of testing this hypothesis. Table 16 also shows a moderate to strong negative correlation holds between the RACE and SES composition variables (%BLACK and %LUNCH) and the PARENT INVOLVEMENT, ORDERLY ENVIRONMENT, and MONITORING & FEEDBACK scales. This would suggest that as the percentage of students on the free or reduced price lunch program increases, These school characteristics are perceived by teachers to get worse.

Correlations between the Scales and Achievement Outcomes

Finally, we begin to ask questions about the relationship between the school characteristics as measured by the Connecticut School Effectiveness Questionnaire (CSEQ) and achievement as measured by the

BSAP exams. Table 17 provides results of a simple bivariate correlation at the school level of analysis. When viewing these findings, the reader should keep in mind that correlations at the school level are always higher than those at the individual level (see Cooley, Bond & Mao, 1981). Interestingly, this analysis shows that only PRINCIPAL LEADERSHIP (which parallels the Connecticut "Instructional Leadership" scale) is not significantly correlated with either outcome measure. PARENT INVOLVEMENT exhibits the strongest relationship with the dependent variables producing correlation coefficients of r = .64 with the reading measure and r = .48 with the math outcome measure. The reader is reminded once again that these are simple bivariate correlations that suffer from inherent "level of analysis" problems. Therefore we should exercise caution in interpreting the findings given thus far. It will be interesting to see if these simple analyses are born out by the final regression models derived using HLM.

HLM Analyses

The central focus of this study involves the use of Hierarchical Linear Modeling to examine the effects of school characteristics on achievement levels and on regression slopes expressing effects of student background factors on achievement. The basic steps of the HLM analyses were outlined in Chapter 3. In this section, the four stages discussed in that chapter will be illustrated in terms of the actual data from South Carolina. In an effort to simplify the presentation,

Table 17

Bivariate Correlations between Final Scales and Outcome Measures.

	Mean Reading Scale Scores	Mean Math Scale Scores
ORDERLY ENVIRONMENT	.23**	.20*
INSTRUCTIONAL OBJECTIVES	.22*	.25**
PRINCIPAL LEADERSHIP	.03	.07
MONITORING & FEEDBACK	.28**	.25**
PARENT INVOLVEMENT	.64**	.48**

Note. All correlations are at the school level (n = 103).

*.01 < p < .05

**One-tailed significance p-value < .01

results of the analyses with reading achievement on the BSAP tests will be given first. The four phases will then be repeated for the math outcome. Given the complexity of the analysis, several approaches could be used to reach the "theoretical" models discussed in phase four. The goal of this analysis is to test the effective schools theory as defined in Chapter 1. At the same time we must guard against specification errors in building these models. The "four phase" approach is designed to minimize (not eliminate) specification errors by gradually building up to the final explanatory models.

Apportioning Parameter Variance for the Reading Outcome

Using the simple model represented by equations [3] and [4] of Chapter 3, we can derive several important pieces of information regarding the reading outcome measure used in this study. First, using the formula for intraclass correlation, we find an estimate for the amount of reading test scale score variation that lies between schools. Both r and σ^2 are provided by the HLM output. The intraclass correlation, ρ_I, was determined to be .085 for the reading achievement outcome. So, with no control for any covariates, about 8.5% of the total variance in student reading achievement is found between schools. Conversely, over 91% of the variance in this outcome measure is potentially explainable by within-school factors alone. This is consistent with the findings of Jencks et al. (1972) reported in Chapter 2, and numerous other studies using similar outcome measures.

The second piece of information referred to above involves a test of the significance of this between school variation where the null

hypothesis is represented symbolically by H_0: $r = 0$. The Chi-square

test provided by HLM indicates that we can reject this null hypothesis

and infer that there are significant differences among schools in their

mean reading achievement outcomes ($\chi^2 = 2335.0$, df $= 102$,

p < .0005).

Finally, HLM provides an estimate of the reliability (w_j) of the

school _sample means_ as estimators of their _true_ means. The point

estimate of this reliability for the reading outcome measure is .944

indicating that there are reliable differences among school means with

regard to the reading outcome measure.

Assessing Homogeneity of Regression for Reading Outcomes

At this stage of the analysis only student level variables are

added to the within school model as depicted in equation [6] of Chapter

3. The variable representing gifted status (i.e., GIFTED) was dropped

from any further analysis because its inclusion leads to a situation

where the derivation of parameter estimates becomes impossible (i.e.,

the EM algorithm could not create an inverted matrix). This leaves us

with the regression model described in Figure 3 on the following page.

Table 18 shows the result of the analysis for this "unconditional"

model. We would predict from the bivariate correlations of the

previous section that the student level indicator of socioeconomic

status (i.e., LUNCH) and ethnic group affiliation (RACE) would have the

strongest effect on mean reading scores, and this prediction is born

out here, as indicated by the t-statistic which is greatest for LUNCH

(t $= 18.1$) and smallest for REPEAT (t $= 10.2$).

All five student-level variables are shown to have a statistically

Figure 3

Unconditional Regression Models for Reading Achievement Scores.

Within-School Model:

$$RSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + \beta_{j2}LUNCH + \beta_{j3}SEX + \beta_{j4}RACE +$$

$$\beta_{j5}REPEAT + R_{ij}$$

Between-School Models:

INTERCEPT $\qquad \beta_{j0} = \gamma_{00} + U_{j0}$

YEAR $\qquad \beta_{j1} = \gamma_{10} + U_{j1}$

LUNCH $\qquad \beta_{j2} = \gamma_{20} + U_{j2}$

SEX $\qquad \beta_{j3} = \gamma_{30} + U_{j3}$

RACE $\qquad \beta_{j4} = \gamma_{40} + U_{j4}$

REPEAT $\qquad \beta_{j5} = \gamma_{50} + U_{j5}$

Note. γ_{00} is the grand mean of the outcome measure RSS_{ij}.
The γ_{jk}'s are the mean within-school regression
coefficients for variable k across all schools in the
sample (k = 1, 2, 3, 4, 5).

Table 18

Unconditional Model with Reading Scale Scores as the Outcome.

	Gamma[a]	Standard Error	t-statistic	p-value
for INTERCEPT (β_{j0})				
γ_{00} BASE	782.28	2.59	301.937	0.000
for YEAR (β_{j1})				
γ_{10} BASE	14.73	1.08	13.661	0.000
for LUNCH (β_{j2})				
γ_{20} BASE	34.01	1.86	18.290	0.000
for SEX (β_{j3})				
γ_{30} BASE	12.14	1.10	11.069	0.000
for RACE (β_{j4})				
γ_{40} BASE	31.76	2.06	15.391	0.000
for REPEAT (β_{j5})				
γ_{50} BASE	30.56	3.00	10.190	0.000

Note. All independent within-school variables are treated as random effects, and (with the exception of YEAR) have been centered on the sample mean value.

[a] Gamma's in this table represent the average slope across all schools (i.e., BASE level) for each of the within-school regression coefficients in Figure 3.

significant effect on mean reading achievement (p < .0005). For example, these results indicate that girls, on average, score 12.14 points higher than boys on the third grade reading test. Also, students <u>not</u> on the lunch program score 34 points higher on the reading exam than do their lower SES counterparts (see γ_{20} in Table 18). Black students perform less well than other ethnic groups by almost 32 scale score units (all else held constant), and non-repeaters outperform students who have been held in grade by 30.56 points, all else held constant.

Table 19 shows that, here again, the mean achievement estimates are very reliable (w_j = .951). This means that over 95% of the observed variation in the mean achievement levels is potentially explainable by specifying school characteristics as predictors. The regression coefficients associated with the student background variables are far less reliable, however. For example, the school-level reliability of .368 for β_{j2} indicates that over 63% of the observed variation in these estimated β_{j2}'s is random error. Table 19 shows the reliability estimates for slopes range from .105 for SEX to .588 for YEAR.

Table 19 also provides evidence that the relationship between student background and achievement varies as a function of what school the students happen to be enrolled in. Said another way, school characteristics may be mediating the relationship between YEAR, LUNCH, SEX, RACE, REPEAT, and reading achievement. Notice that, with the exception of the SEX slope, Chi-square tests of variation between school slope estimates are significant at the .001 level. The task now

Table 19

Chi-Square Table for the Unconditional RSS Model.

Parameter		Var (β_{jk})	d.f.	Chi-square	p-value	w_j
β_{j0}	INTERCEPT	658.288	102	2660.6	0.000	.951
β_{j1}	YEAR SLOPE	74.860	102	291.14	0.000	.588
β_{j2}	LUNCH SLOPE	163.542	102	193.66	0.000	.368
β_{j3}	SEX SLOPE	16.480	102	134.38	0.017	.105
β_{j4}	RACE SLOPE	205.172	102	191.31	0.000	.253
β_{j5}	REPEAT SLOPE	344.419	102	154.09	0.001	.246

Note. Var (β_{jk}) = estimated parameter variance. The reliability estimate (w_j) represents the proportion of variance in the OLS within-school estimates that is parameter variance (Bryk et al., 1986). See the discussion regarding equation [5] in Chapter 3 for more information.

is to determine which school characteristics may be influencing these between-school differences.

Testing for School Composition Effects on Reading Achievement

In the third general stage of the HLM analysis, the between-school model of mean reading achievement is extended to account for the effects of school composition variables. This model is described in Figure 4. Recall, that the variables in the between-school portion of this model represent the percentage of students in each school that were classified into specific subgroups based on their individual background measures.

Table 20 shows the estimates of the between-school model for reading scale scores. The table indicates that, both %LUNCH and %GIFTED are significantly related to third grade mean reading achievement in these schools (t-values are -4.615 and 2.792 respectively). This result informs us that as the percentage of students on the free and reduced price lunch program increases, the RSS for a "typical" student in that school will decrease (all other factors in both the between- and within-school portion of the model held constant). The opposite effect holds for %GIFTED. Before proceeding, the term "typical" may need some clarification. In the sense used here, a typical student receives the value of zero on all background variables. Of course this is impossible (as the code book of Appendix A shows). This reference to a typical student (or school) then is just a convenient way of interpreting results, and nothing more should be inferred.

The careful observer might wonder why %BLACK does not appear to

Figure 4

School Composition Models for Reading Achievement Scores.

Within-School Model:

$$RSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + \beta_{j2}LUNCH + \beta_{j3}SEX + \beta_{j4}RACE +$$
$$\beta_{j5}REPEAT + R_{ij}$$

Between-School Models:

INTERCEPT $\quad \beta_{j0} = \gamma_{00} + \gamma_{01}\%LUNCH + \gamma_{02}\%BLACK +$

$$\gamma_{03}\%REPEAT + \gamma_{04}\%GIFTED + \gamma_{05}SIZE + U_{j0}$$

YEAR $\quad \beta_{j1} = \gamma_{10} + U_{j1}$

LUNCH $\quad \beta_{j2} = \gamma_{20} + \gamma_{21}\%LUNCH + U_{j2}$

SEX $\quad \beta_{j3} = \gamma_{30} + U_{j3}$

RACE $\quad \beta_{j4} = \gamma_{40} + \gamma_{41}\%BLACK + U_{j4}$

REPEAT $\quad \beta_{j5} = \gamma_{50} + \gamma_{51}\%REPEAT + U_{j5}$

Note. γ_{00} is the grand mean of the outcome measure RSS_{ij}.

Table 20

School Composition Model for the Reading Outcome.

		Gamma	Standard Error	t-statistic	p-value
for INTERCEPT (β_{j0})					
γ_{00}	BASE	782.110	1.787	437.653	0.000
γ_{01}	%LUNCH	-0.605	0.131	-4.615	0.000
γ_{02}	%BLACK	-0.125	0.112	-1.116	0.265
γ_{03}	%REPEAT	-0.426	0.407	-1.047	0.296
γ_{04}	%GIFTED	0.708	0.254	2.792	0.006
γ_{05}	SIZE	-0.249	0.170	-1.464	0.143
for YEAR (β_{j1})					
γ_{10}	BASE	14.647	1.080	13.562	0.000
for LUNCH (β_{j2})					
γ_{20}	BASE	33.723	1.858	18.154	0.000
γ_{21}	%LUNCH	-0.141	0.085	-1.668	0.095
for SEX (β_{j3})					
γ_{30}	BASE	12.151	1.126	10.791	0.000
for RACE (β_{j4})					
γ_{40}	BASE	31.769	2.081	15.263	0.000
γ_{41}	%BLACK	-0.067	0.092	-0.729	0.466
for REPEAT (β_{j5})					
γ_{50}	BASE	33.548	2.858	11.737	0.000
γ_{51}	%REPEAT	-2.085	0.555	-3.757	0.000

Note. All school-level composition variables, and SIZE, have been centered on their respective sample means. The BASE for each β_{jk} represents the mean value for that within-school regression coefficient.

have a significant impact on the INTERCEPT (β_{j0}) although it was more highly correlated with both achievement measures than was %GIFTED in bivariate analyses. By regressing the Empirical Bayes (EB) residuals on our composition variables during the HLM analysis of the unconditional model we found the t-value for %LUNCH, when it alone is used to model the INTERCEPT, was -10.6. Corresponding values for %BLACK and %GIFTED were t = -8.1 and t = 3.9 respectively. This indicates that the high collinearity between %LUNCH and %BLACK at the school level (r = .87) is causing problems in estimating the strength of this relationship. Finally, notice that relevant composition variables are used to model β_{j2}, β_{j4}, and β_{j5}. This was done in an attempt to account for possible grouping effects.

Assessing the Effects of School Characteristics on Reading Scores

We are now in a position to address the final two research questions dealing with the influence of school effectiveness characteristics on the average achievement levels in schools (i.e, _quality_ of schooling), and the effects these of school characteristics on the effects of individual background variables (i.e., the _equity_ issue). The final model, depicted in Figure 5, reflects effective schools theory in that we are using all five school characteristics to model the INTERCEPT and the slopes of our important background variables (i.e., RACE and LUNCH).

Notice that %LUNCH, %REPEAT, and %GIFTED also remain in the between-schools equation to model the INTERCEPT, and that only %REPEAT is used to model the REPEAT slope. Other composition variables that did not have a significant "group effect" on their respective slopes

Figure 5

Final Theoretical Model for Reading Achievement Scores.

Within-School Model:

$$RSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + \beta_{j2}LUNCH + \beta_{j3}SEX + \beta_{j4}RACE +$$

$$\beta_{j5}REPEAT + R_{ij}$$

Between-School Model:

for INTERCEPT $\quad \beta_{j0} = \gamma_{00} + \gamma_{01}\%LUNCH + \gamma_{02}\%REPEAT +$

$$\gamma_{03}\%GIFTED + \gamma_{04}MONITORING \text{ \& } FEEDBACK +$$

$$\gamma_{05}PRINCIPAL \text{ } LEADERSHIP + \gamma_{06}INSTRUCTIONAL \text{ } OBJECTIVES +$$

$$\gamma_{07}PARENT \text{ } INVOLVEMENT + \gamma_{08}ORDERLY \text{ } ENVIRONMENT + U_{j0}$$

for YEAR $\quad \beta_{j1} = \gamma_{10} + U_{j1}$

for LUNCH $\quad \beta_{j2} = \gamma_{20} + \gamma_{21}MONITORING \text{ \& } FEEDBACK +$

$$\gamma_{22}PRINCIPAL \text{ } LEADERSHIP + \gamma_{23}INSTRUCTIONAL \text{ } OBJECTIVES +$$

$$\gamma_{24}PARENT \text{ } INVOLVEMENT + \gamma_{25}ORDERLY \text{ } ENVIRONMENT + U_{j2}$$

for SEX $\quad \beta_{j3} = \gamma_{30} + U_{j3}$

for RACE $\quad \beta_{j4} = \gamma_{40} + \gamma_{41}MONITORING \text{ \& } FEEDBACK +$

$$\gamma_{42}PRINCIPAL \text{ } LEADERSHIP + \gamma_{43}INSTRUCTIONAL \text{ } OBJECTIVES +$$

$$\gamma_{44}PARENT \text{ } INVOLVEMENT + \gamma_{45}ORDERLY \text{ } ENVIRONMENT + U_{j4}$$

for REPEAT $\quad \beta_{j5} = \gamma_{50} + \gamma_{51}\%REPEAT + U_{j5}$

___

Note. γ_{00} is the grand mean of the outcome measure RSS_{ij}, and the γ_{jk}'s are the mean within-school regression coefficients for variable k across all schools in the sample (k = 1, 2, 3, 4, 5).

(including %BLACK) were deleted from the model. The results of this analysis are displayed in Table 21.

Table 21 shows that %LUNCH and %GIFTED are still related to the intercept, and %REPEAT has a significant effect on the slope coefficient for REPEAT (β_{j5}). That is, as the proportion of repeaters in a school increases, the achievement gap between repeaters and students not held in grade narrows. The negative effect of %LUNCH on the INTERCEPT indicates that as the percentage of students on the free or reduced price lunch program in a school increases, the average reading achievement for the school will decrease (all else held constant). As we would expect %GIFTED has the opposite effect.

But, what effects are the school effectiveness characteristics having on either the INTERCEPT or slope coefficients for the background variables? Table 21 shows that PARENT INVOLVEMENT has a statistically significant effect (t - 2.369) on the average reading scale score. In this model we test five a-priori hypotheses using Dunn's multiple comparison procedure as described by Kirk (1982). If we set the familywise error rate at α - .05, with 5 planned contrasts and 102 degrees of freedom, the t-value must exceed 2.364 before we can reject the hypothesis of no effect (see Dayton & Schafer, 1973, Table 1).[10] Table 21 shows that the t-statistic for the remaining four characteristics all fall short of this critical value. Therefore, we

---

[10] The one-tailed value at α - .01, with 100 degrees of freedom, is 2.946 (from Dayton & Schafer, 1973, Table 2).

Table 21

Gamma Table for the Reading Scale Score Theoretical Model.

	Gamma	Standard Error	t-statistic	
for INTERCEPT (β_{j0})				
γ_{00} BASE	781.971	1.713	456.572	
γ_{01} %LUNCH	-0.485	0.108	-4.484	**
γ_{02} %REPEAT	-0.262	0.404	-0.648	
γ_{03} %GIFTED	0.689	0.252	2.733	*
γ_{04} MONITORING & FEEDBACK	9.904	14.989	0.661	
γ_{05} PRINCIPAL LEADERSHIP	-11.305	6.051	-1.868	
γ_{06} INSTRUCTIONAL OBJECTIVES	0.855	10.018	0.085	
γ_{07} PARENT INVOLVEMENT	15.226	6.428	2.369	*
γ_{08} ORDERLY ENVIRONMENT	-3.347	4.793	-0.689	
for YEAR (β_{j1})				
γ_{10} BASE	14.621	1.082	13.510	
for LUNCH (β_{j2})				
γ_{20} BASE	33.895	1.773	19.118	
γ_{21} MONITORING & FEEDBACK	11.928	15.501	0.769	
γ_{22} PRINCIPAL LEADERSHIP	7.956	6.158	1.292	
γ_{23} INSTRUCTIONAL OBJECTIVES	-26.832	10.155	-2.642	*
γ_{24} PARENT INVOLVEMENT	9.205	5.302	1.736	
γ_{25} ORDERLY ENVIRONMENT	4.309	5.135	0.839	
for SEX (β_{j3})				
γ_{30} BASE	11.991	1.124	10.673	
for RACE (β_{j4})				
γ_{40} BASE	32.134	2.082	15.438	
γ_{41} MONITORING & FEEDBACK	8.543	17.597	0.485	
γ_{42} PRINCIPAL LEADERSHIP	6.180	6.890	0.897	
γ_{43} INSTRUCTIONAL OBJECTIVES	-3.996	11.765	-0.340	
γ_{44} PARENT INVOLVEMENT	11.112	5.652	-1.966	
γ_{45} ORDERLY ENVIRONMENT	-7.903	5.833	-1.355	
for REPEAT (β_{j5})				
γ_{50} BASE	33.215	2.857	11.625	
γ_{51} %REPEAT	-2.145	0.555	-3.862	

Note. Table entries correspond to Figure 5.

   * Family Type I error risk for directional test < .05
   ** Family Type I error risk for directional test < .01

retain the null hypothesis for MONITORING & FEEDBACK, PRINCIPAL

LEADERSHIP, INSTRUCTIONAL OBJECTIVES, and ORDERLY ENVIRONMENT and

conclude that only PARENT INVOLVEMENT has a significant effect on

average third grade reading achievement scores. Centering all student-

and school-level variables in this model allows a direct interpretation

of the results. That is, a one unit increase in a typical school's

rating on the PARENT INVOLVEMENT scale adds about 15.2 points to the

average achievement for that school.

The INSTRUCTIONAL OBJECTIVES variable appears to have an effect on

the slope for our proxy measure of socioeconomic status (i.e., β_{j2}).

This result provides us with an example of the effect this

characteristic has on equity in schooling. To help interpret Table 21,

we need to present some basic statistics for the variables involved in

the model outlined in Figure 5. Table 22 provides information

regarding the school level means and standard deviations for the five

effective schools characteristics and important composition variables

plus the variable SIZE which is a rough indicator of the number of

teachers working in each school.[11] Data are also given for the

minimum and maximum values of these variables for all schools in the

HLM analysis.

For purposes of illustration, we will refer to the general model

---

[11]   More accurately, SIZE represents the number of staff members who
       returned a completed CSEQ form to the University of South
       Carolina School Council Assistance Project office.

Table 22

Basic Statistics Describing School Characteristics.

Variable	MEAN[a]	SD	MIN	MAX
MONITORING & FEEDBACK	4.11	.26	2.77	4.64
PRINCIPAL LEADERSHIP	3.75	.43	2.03	4.60
INSTRUCTIONAL OBJECTIVES	4.06	.35	1.99	4.70
PARENT INVOLVEMENT	3.73	.40	2.32	4.17
ORDERLY ENVIRONMENT	3.82	.47	2.62	4.58
%LUNCH	52.83	24.00	5.9	97.8
%GIFTED	4.90	6.68	0.0	33.3
%REPEAT	6.49	4.55	0.2	21.8
SIZE	28.02	10.37	9	58

Note. The variable values used in all HLM models have been deviated from these sample means (i.e., they have been centered around the sample mean).

[a] Mean of school means (n = 103).

illustrated in Figure 5. Substituting the right side of the between-school model for LUNCH into β_{j2} of the within-school model, we get the following equation:

$$RSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + [\; \gamma_{20} + \gamma_{21}MONITORING \; \& \; FEEDBACK +$$
$$\gamma_{22}PRINCIPAL \; LEADERSHIP + \gamma_{23}INSTRUCTIONAL \; OBJECTIVES +$$
$$\gamma_{24}PARENT \; INVOLVEMENT + \gamma_{25}ORDERLY \; ENVIRONMENT + U_{j2} \;]LUNCH +$$
$$\beta_{j3}SEX + \beta_{j4}RACE + \beta_{j5}REPEAT + R_{ij}$$

This formula allows us to compare RSS for low SES students (i.e., those for whom LUNCH = -.507) and other children (where LUNCH = .493) across schools in our sample. Further, if we restrict the discussion to "typical" students (i.e., students for whom SEX, RACE, AND REPEAT all equal zero), and assume measurement errors are indeed random (i.e., U_{j2} and R_{ij} equal zero) we arrive at the following reduced equation:

$$\hat{RSS} = \beta_{j0} + [\; \gamma_{20} + \gamma_{21}MONITORING \; \& \; FEEDBACK +$$
$$\gamma_{22}PRINCIPAL \; LEADERSHIP + \gamma_{23}INSTRUCTIONAL \; OBJECTIVES +$$
$$\gamma_{24}PARENT \; INVOLVEMENT + \gamma_{25}ORDERLY \; ENVIRONMENT]LUNCH$$

Using values from Tables 21 and 22, we can now compute the expected reading scale scores for "typical" students in these schools. In the above equation MONITORING & FEEDBACK, PRINCIPAL LEADERSHIP, PARENT INVOLVEMENT, and ORDERLY ENVIRONMENT are replaced by their mean values and multiplied by their respective Gamma's. INSTRUCTIONAL OBJECTIVES, for the purposes of this example, was allowed to vary by plus or minus

one standard deviation from its mean yielding a low value of -.35 and a high of +.35. Figure 6 provides a graphical representation for the result of our calculations. It is clear from this illustration that an interaction does exist between our proxy for SES and the INSTRUCTIONAL OBJECTIVES scale. We see, with the aid of Figure 6, that the SES gap in reading achievement narrows from 43.3 to 24.9 scale score points as the schools' original rating on INSTRUCTIONAL OBJECTIVES increases from 3.71 to 4.41. Unfortunately, this "equalizing" effect comes about, in part, because higher SES students, on average, perform less well on the BSAP exam when school-wide objectives become the focal point of instruction. That is, as lower SES students gain 10.1 points on the reading achievement scale, higher SES students lose 8.6 scale score points across this range of values for INSTRUCTIONAL OBJECTIVES.

Finally, Table 23 contains the Chi-square values and parameter reliability estimates for the "theoretical" reading scale score model. Table 24 shows the changes in estimated parameter variance from the unconditional model to the final theoretical model for the INTERCEPT in the between-schools model and for RACE and LUNCH slopes as estimated by HLM. The percentages in Table 24 are derived with the following formula

$$\%R^2 = [[Var\ (\beta_u) - Var\ (\beta_h)]/\ Var\ (\beta_u)]\ X\ 100$$

where Var (β_u) represents the estimated parameter variance of the unconditional model, and Var (β_h) is the estimated parameter variance for the composition or theoretical model.

As can be seen, the estimated parameter variance actually increases

Figure 6

Interaction between INSTRUCTIONAL OBJECTIVES and LUNCH in RSS Model.
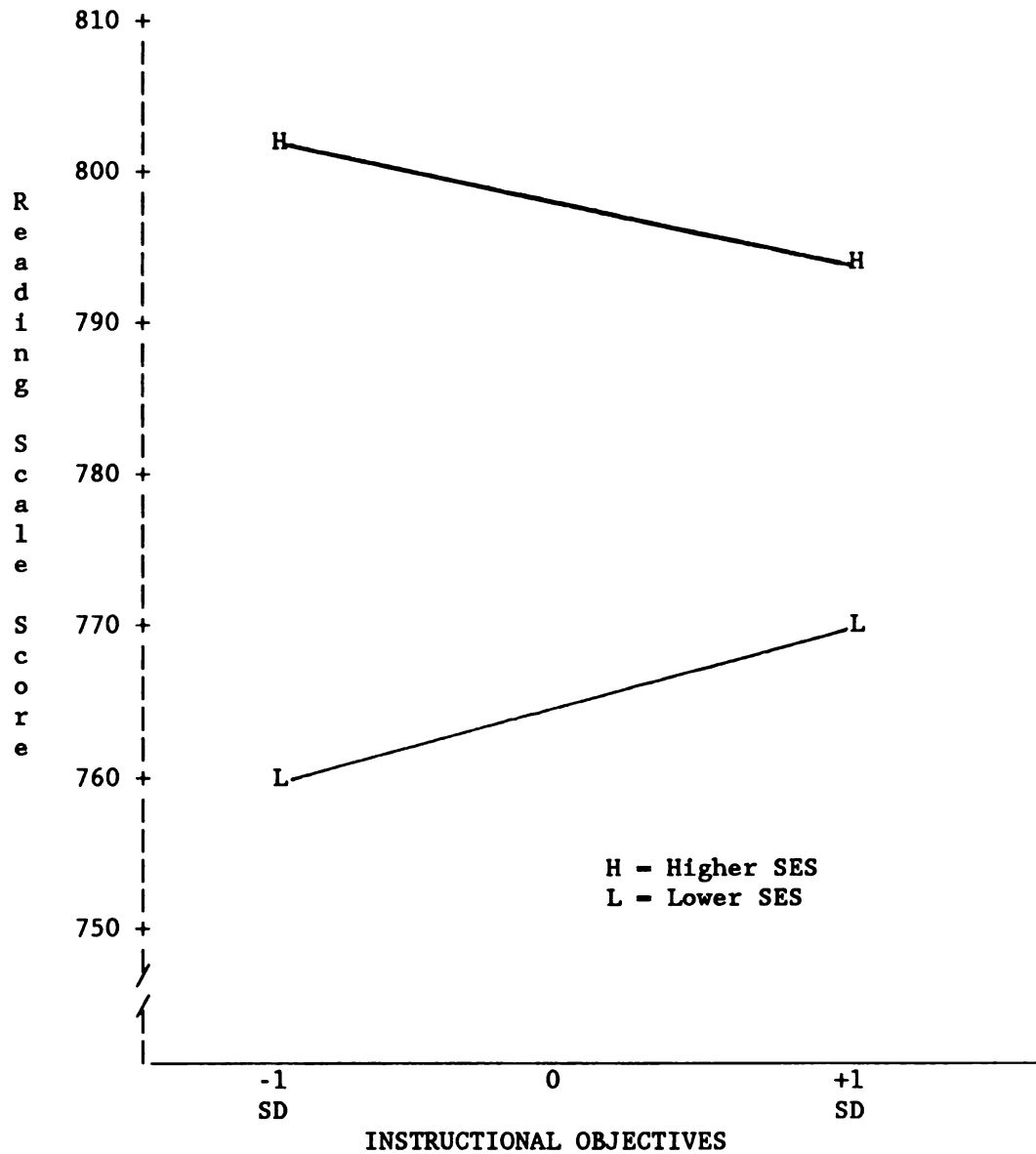
Note. The range of values represented by ±1 SD is 3.71 to 4.41

Table 23

Chi-Square Table for Theoretical Reading Model.

	Parameter		Var (β_{jk})	d.f.	Chi-square	p-value	w_j
β_{j0}	INTERCEPT		269.503	94	1013.3	0.000	.888
β_{j1}	YEAR	SLOPE	75.671	102	291.24	0.000	.590
β_{j2}	LUNCH	SLOPE	130.926	97	171.66	0.000	.318
β_{j3}	SEX	SLOPE	22.298	102	134.44	0.017	.137
β_{j4}	RACE	SLOPE	208.306	97	184.15	0.000	.256
β_{j5}	REPEAT	SLOPE	191.410	101	132.86	0.018	.154

Note. All values in this tabel were drawn from the HLM output.

Table 24

Variance Accounted for by Reading Achievement Models.

Estimated Parameter Variance

Model	INTERCEPT	LUNCH	RACE
UNCONDITIONAL	658.288	163.542	205.172
COMPOSITION	295.935 (55.0)	156.613 (4.2)	206.053 (-0.0)
THEORETICAL	269.503 (59.1)	130.926 (18.7)	208.306 (-1.5)

Note. The number in parentheses represents the percentage of variance accounted for by the HLM model (using the UNCONDITIONAL model as a baseline).

slightly for RACE in the theoretical model from what it was in the unconditional model. This result indicates that our theory falls short in terms of modeling the RACE parameter. The negative value for RACE (i.e., -1.5%) indicates that less estimated parameter variance for β_{j4} is explained by the theoretical model than was true for the unconditional model. This may result from specification errors.

For the INTERCEPT, explained variance is substantial relative to total estimated parameter variance. We should recognized however, that school composition variables (e.g., percentage of students on the free lunch program), and not our five effective schools characteristics, account for the largest change in estimated parameter variance for mean reading achievement. For example, the COMPOSITION model reduced the estimated parameter variance for the INTERCEPT from 658.3 to 295.9. This implies that the school composition variables _alone_ account for roughly 55% of the explainable variance after controlling for student-level background variables. Adding our measures of the five school characteristics to this model explains another 4.1% of the variance in β_{j0} for the reading model.

The situation with respect to the LUNCH slope is very different. Here the effective schools characteristics account for the lion's share of the explainable parameter variance (i.e., 14.5% above that accounted for by the composition model). The reliability estimates (w_j's) for these differentiation effects, shown in Table 23, give the reader a sense of how much of the variability among a set of regression coefficients is likely to be explainable by our school-level variables. Here nearly 89% of the observed variation in reading

achievement levels is potentially explainable by specifying school-level practices, policies, and/or characteristics as predictors.

Apportioning Parameter Variance for the Math Outcome

The procedures used to address the quality and equity research questions for mathematics achievement are nearly identical to those discussed for reading. First, we can examine the results of the simplest HLM model represented by equations [3] and [4] in Chapter 3 with the math scale score (MSS) as the outcome of interest. Here the intraclass correlation is computed to be ρ_I = .107 indicating that nearly 11% of the total variance in math achievement is between schools. We would expect this value to be larger than that obtained for reading scale scores under the assumption that relatively more math instruction takes place in the school than in the home.

The point reliability estimate for this outcome measure is .957 which reveals that the differences among school means on the math outcome are reliable estimators of the true school means. Once again a Chi-square of 2945.5 with 102 degrees of freedom allows us to infer that there are significant differences among schools with regard to their mean mathematics achievement outcomes (p < .0005).

Effects of Individual Level Variables on Math Achievement

In phase two we create the unconditional model (see Figure 7) to study the impact of student-level variables on Math Scale Scores (MSS). The result of this HLM run is shown in Table 25. Two differences from the RSS unconditional model for reading achievement deserve further mention. First, notice that Table 25 contains no

Figure 7

Unconditional Regression Models for Mathematics Achievement Scores.

Within-School Model:

$$MSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + \beta_{j2}LUNCH + \beta_{j3}RACE +$$
$$\beta_{j4}REPEAT + M_{ij}$$

Between-School Models:

INTERCEPT $\qquad \beta_{j0} = \gamma_{00} + V_{j0}$

YEAR $\qquad \beta_{j1} = \gamma_{10} + V_{j1}$

LUNCH $\qquad \beta_{j2} = \gamma_{20} + V_{j2}$

RACE $\qquad \beta_{j3} = \gamma_{30} + V_{j3}$

REPEAT $\qquad \beta_{j4} = \gamma_{40} + V_{j4}$

Note. γ_{00} is the grand mean of the outcome measure MSS_{ij}.

Table 25

Unconditional Model with MSS as the Outcome.

	Gamma	Standard Error	t-statistic	p-value
for INTERCEPT (β_{j0})				
γ_{00} BASE	787.778	3.589	219.471	0.000
for YEAR (β_{j1})				
γ_{10} BASE	-7.453	2.026	-3.680	0.000
for LUNCH (β_{j2})				
γ_{20} BASE	29.718	2.180	13.632	0.000
for RACE (β_{j3})				
γ_{30} BASE	39.645	2.399	16.524	0.000
for REPEAT (β_{j4})				
γ_{40} BASE	17.616	2.698	6.528	0.000

Note. The residual variance for REPEAT was set to zero. Parameter labels correspond to Figure 7.

entry for SEX. This variable was found to have no significant effect on individual math achievement and was therefore dropped from the unconditional MSS model. Also notice that the YEAR variable has a negative relationship with the MSS outcome. This result was predicted in our discussion of Table 5. That is, as a group, the students who took the BSAP math exams in 1986 did less well than their third grade predecessors in 1984. The remaining within-school regression slopes can be interpreted as before. For example, students not on the school lunch program, as a group, score almost 30 points higher on the BSAP math exam than do their lower SES classmates, and Blacks score about 40 points below other ethnic groups, on average.

Finally, Table 26 provides evidence that the effects of YEAR, LUNCH, and RACE differ across schools. Notice that REPEAT does not appear in this table. While the variable REPEAT had a statistically significant effect (p < .01) on the outcome measure, that effect does not vary significantly across schools. Therefore, REPEAT is treated as a fixed effect in this model (i.e., its residual variance was set to zero).

Effects of Composition Variables on Math Achievement

Figure 8 shows the school-level composition variables that were added to the unconditional model in this third phase of the HLM analysis. As before, all school-level variables not associated with effective schools research per se are used to model the INTERCEPT (β_{j0}) of the within-school regression model. Also the within-schools slopes for LUNCH, RACE, and REPEAT are modeled with their associated composition variables (i.e., %LUNCH, %BLACK, and %REPEAT respectively).

Table 26

Chi-Square Table for the Unconditional Math Scale Score Model.

Parameter		Var (β_{jk})	d.f.	Chi-square	p-value	w_j
β_{j0} INTERCEPT		1276.619	102	3297.4	0.000	.961
β_{j1} YEAR	SLOPE	350.641	102	661.85	0.000	.817
β_{j2} LUNCH	SLOPE	205.984	102	165.33	0.000	.329
β_{j3} RACE	SLOPE	246.381	102	164.31	0.000	.213

Note. Entries in this table are derived from HLM output.

Figure 8

<u>School Composition Model for Mathematics Achievement Scores</u>.

Within-School Model:

$$RSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + \beta_{j2}LUNCH + \beta_{j3}RACE +$$
$$\beta_{j4}REPEAT + M_{ij}$$

Between-School Models:

INTERCEPT $\quad \beta_{j0} = \gamma_{00} + \gamma_{01}\%LUNCH + \gamma_{02}\%BLACK + \gamma_{03}\%REPEAT +$
$$\gamma_{04}\%GIFTED + \gamma_{05}SIZE + V_{j0}$$

YEAR $\quad \beta_{j1} = \gamma_{10} + V_{j1}$

LUNCH $\quad \beta_{j2} = \gamma_{20} + \gamma_{21}\%LUNCH + V_{j2}$

RACE $\quad \beta_{j3} = \gamma_{30} + \gamma_{31}\%BLACK + V_{j3}$

REPEAT $\quad \beta_{j4} = \gamma_{40} + \gamma_{41}\%REPEAT + V_{j4}$

Table 27 shows that the results for the MSS model are again slightly different from those for the reading outcome. Now we notice that %LUNCH and SIZE have a statistically significant effect on the INTERCEPT (β_{j0}). The composition variable %LUNCH has the same general effect on math achievement as it did for reading. That is, the more students at a school on the free lunch program, the lower the average achievement score will be. The results in Table 27 also indicate that larger schools tend to produce lower average achievement.

Interestingly, though at the bivariate level %LUNCH and %BLACK are correlated positively with each other (r - .87), their effects on the INTERCEPT $(\gamma_{01}$ - -.842 and γ_{02} - .084 respectively) appear to act in opposite directions. This again illustrates the multicollinearity problem emerging from the findings for our reading scale score model. This result is common in multiple regression analysis where two predictors are highly correlated. The collinearity problem was once again addressed by dropping %BLACK from the model, and retaining %LUNCH and SIZE as we move to the last phase of the analysis.[12]

Effects of School Characteristics on Math Achievement

Again effective schools theory dictates that we use all five measures of our school characteristics to model the INTERCEPT, LUNCH, and RACE parameters as shown in Figure 9. Table 28 shows that no

---

[12] The decision to retain %LUNCH instead of %BLACK was further supported by the fact that the former variable also has a significant effect on the LUNCH slope (β_{j2}).

Table 27

Model with School Composition Variables for MSS Outcome.

		Gamma	Standard Error	t-statistic	p-value
for INTERCEPT (β_{j0})					
γ_{00}	BASE	787.730	3.069	256.708	0.000
γ_{01}	%LUNCH	-0.842	0.231	-3.643	0.000
γ_{02}	%BLACK	0.084	0.195	0.429	0.667
γ_{03}	%REPEAT	-0.714	0.682	-1.048	0.295
γ_{04}	%GIFTED	0.299	0.447	0.670	0.503
γ_{05}	SIZE	-0.719	0.299	-2.400	0.017
for YEAR (β_{j1})					
γ_{10}	BASE	-7.490	2.026	-3.698	0.000
for LUNCH (β_{j2})					
γ_{20}	BASE	29.081	2.130	13.650	0.000
γ_{21}	%LUNCH	-0.292	0.098	-2.985	0.003
for RACE (β_{j3})					
γ_{30}	BASE	39.293	2.379	16.518	0.000
γ_{31}	%BLACK	-0.195	0.108	-1.799	0.071
for REPEAT[a] (β_{j4})					
γ_{40}	BASE	17.589	2.698	6.518	0.000

Note.   All within-school variables have been centered on the
sample mean value.

[a] The REPEAT variable was treated as a fixed effect.

Figure 9

<u>Final Theoretical Model for Math Achievement Scores</u>.

Within-School Model:

$$MSS_{ij} = \beta_{j0} + \beta_{j1}YEAR + \beta_{j2}LUNCH + \beta_{j3}RACE + \beta_{j4}REPEAT + M_{ij}$$

Between-School Models:

for INTERCEPT $\quad \beta_{j0} = \gamma_{00} + \gamma_{01}\%LUNCH + \gamma_{02}SIZE +$

$\gamma_{03}MONITORING \ \& \ FEEDBACK + \gamma_{04}PRINCIPAL \ LEADERSHIP +$

$\gamma_{05}INSTRUCTIONAL \ OBJECTIVES + \gamma_{06}PARENT \ INVOLVEMENT +$

$\gamma_{07}ORDERLY \ ENVIRONMENT + V_{j0}$

for YEAR $\quad \beta_{j1} = \gamma_{10} + V_{j1}$

for LUNCH $\quad \beta_{j2} = \gamma_{20} + \gamma_{21}\%LUNCH + \gamma_{22}MONITORING \ \& \ FEEDBACK$

$+ \ \gamma_{23}PRINCIPAL \ LEADERSHIP +$

$\gamma_{24}INSTRUCTIONAL \ OBJECTIVES +$

$\gamma_{25}PARENT \ INVOLVEMENT + \gamma_{26}ORDERLY \ ENVIRONMENT + V_{j2}$

for RACE $\quad \beta_{j3} = \gamma_{30} + \gamma_{31}MONITORING \ \& \ FEEDBACK +$

$\gamma_{32}PRINCIPAL \ LEADERSHIP +$

$\gamma_{33}INSTRUCTIONAL \ OBJECTIVES +$

$\gamma_{34}PARENT \ INVOLVEMENT +$

$\gamma_{35}ORDERLY \ ENVIRONMENT + V_{j3}$

for REPEAT $\quad \beta_{j4} = \gamma_{40} + V_{j4}$

Note: γ_{00} is the grand mean of the outcome measure MSS_{ij}.

Table 28

Theoretical Model for Math Scale Scores.

		Gamma	Standard Error	t-statistic	
for INTERCEPT (β_{j0})					
γ_{00}	BASE	787.692	3.041	259.017	
γ_{01}	%LUNCH	-0.557	0.201	-2.767	**
γ_{02}	SIZE	-0.650	0.299	-2.179	*
γ_{03}	MONITORING & FEEDBACK	-8.071	26.785	-0.301	
γ_{04}	PRINCIPAL LEADERSHIP	-3.396	10.893	-0.312	
γ_{05}	INSTRUCTIONAL OBJECTIVES	5.896	17.425	0.338	
γ_{06}	PARENT INVOLVEMENT	21.969	11.408	1.926	
γ_{07}	ORDERLY ENVIRONMENT	0.183	8.699	0.021	
for YEAR (β_{j1})					
γ_{10}	BASE	-7.499	2.025	-3.703	
for LUNCH (β_{j2})					
γ_{20}	BASE	29.292	2.158	13.574	
γ_{21}	%LUNCH	-0.134	0.150	-0.892	
γ_{22}	MONITORING & FEEDBACK	17.341	20.035	0.866	
γ_{23}	PRINCIPAL LEADERSHIP	4.442	7.880	0.564	
γ_{24}	INSTRUCTIONAL OBJECTIVES	-27.503	12.924	-2.128	
γ_{25}	PARENT INVOLVEMENT	11.637	8.595	1.354	
γ_{26}	ORDERLY ENVIRONMENT	-2.162	6.151	-0.351	
for RACE (β_{j3})					
γ_{30}	BASE	39.203	2.345	16.716	
γ_{31}	MONITORING & FEEDBACK	-18.160	20.284	-0.895	
γ_{32}	PRINCIPAL LEADERSHIP	-5.594	7.920	-0.706	
γ_{33}	INSTRUCTIONAL OBJECTIVES	19.301	13.645	1.415	
γ_{34}	PARENT INVOLVEMENT	19.541	6.485	3.012	**
γ_{35}	ORDERLY ENVIRONMENT	-3.130	6.721	-0.466	
for REPEAT[a] (β_{j4})					
γ_{40}	BASE	17.562	2.698	6.508	

[a]The residual variance for REPEAT slope has been set to zero.

* Family Type I error risk for directional test < .05
** Family Type I error risk for directional test < .01

significant effect is observed for any of the five effective schools

characteristics on average math achievement. However, PARENT

INVOLVEMENT has a definite effect on the slope associated with our RACE

variable (β_{j3}).

With the results from Table 28 and the aid of Figure 9 we can once

again provide a concrete example of how these characteristics are

affecting math achievement in our sample of schools. We first reduce

the model depicted in Figure 9 by assuming YEAR, LUNCH, REPEAT, M_{ij},

and V_{j3} all equal zero. Since the variables of interest here are

RACE and PARENT INVOLVEMENT, only these will be allowed to change in

value in the following equation. We have assigned the sample mean

value to the other variables so our computational formula reduces to:

$$\hat{MSS} = 787.69 + 21.97(PARENT\ INVOLVEMENT) +$$
$$[39.20 + 19.54(PARENT\ INVOLVEMENT)]RACE$$

Now when we substitute the high and low values of PARENT INVOLVEMENT

(i.e., $\pm.40$), and our student-level RACE values (-.575 for Blacks and

.425 for others) in this equation we see that the math achievement

scores of Black students increases by 8.6 points as PARENT INVOLVEMENT

increases from one standard deviation below to one standard deviation

above zero. While this may seem encouraging, the increase for students

of other racial/ethnic groups across this same range of PARENT

INVOLVEMENT values is about 24.2 scale score units. Said another way,

the achievement gap between these two groups (i.e., Blacks and Others)

widens from 31.4 to 47.0 points as the relations with the home are seen

to improve in our sample of schools (see Figure 10). This result is just the opposite of what we would expect to find in an effective school as defined in Chapter 1. It does support our findings regarding the bivariate correlations between these variables, however. We may even venture that the direction of the effect is not what effective schools theory would suggest. That is, both achievement scores _and_ PARENT INVOLVEMENT apparently improve as the percentage of White students in the typical school increases. This implies that PARENT INVOLVEMENT did not precede achievement gains in time. Longitudinal analyses would have to be performed to test this conclusion regarding causal ordering.

Using Table 29 we can look at changes in estimated parameter variance that occur between the unconditional model and the final theoretical model for math scores. We see in Table 30 that the theoretical model explains a sizable portion of the parameter variance in the remaining three variables of interest. The final model reduced the estimated parameter variance for the INTERCEPT from 1276.6 to 901.8 a change of 29.4%. Corresponding changes for LUNCH and RACE slopes are 11.1% and 13.3% respectively. It would appear that the final model provides a better fit for our data than did the unconditional math model. However, the relatively small increase in estimated parameter variance explained when these school level variables are added to the model indicates that we have not yet found the most important variables for modeling school achievement or that we cannot isolate the signal from noise in measuring our constructs.

Figure 10

<u>Interaction between PARENT INVOLVEMENT and RACE in the MSS Model</u>.

PARENT INVOLVEMENT
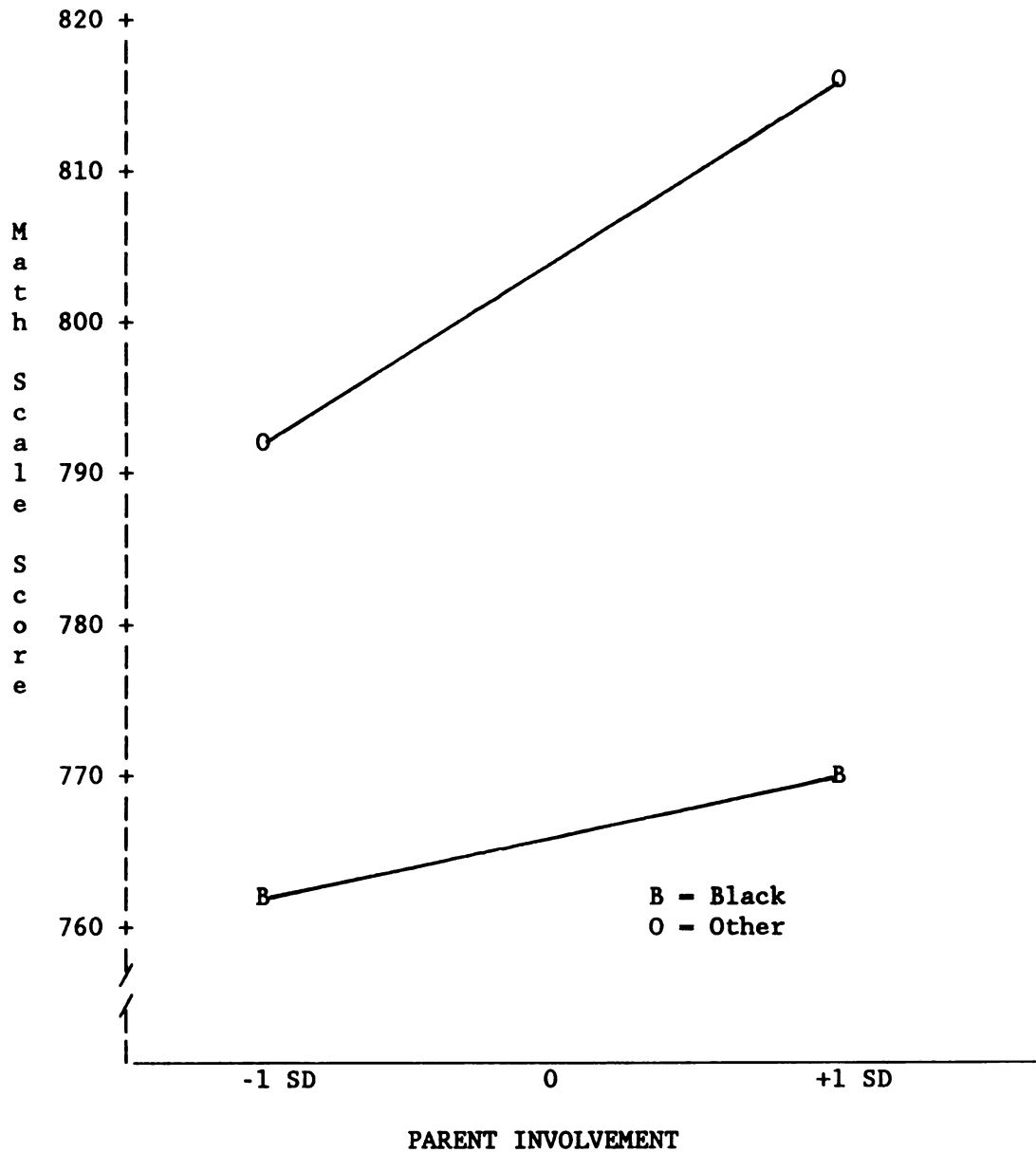
<u>Note</u>. The range of values represented by ±1 SD is 3.33 to 4.13

Table 29

Chi-Square Table for Final Theoretical MSS Model.

	Parameter		Var (β_{jk})	d.f.	Chi-square	p-value	w_j
β_{j0}	INTERCEPT		901.791	95	2167.7	0.000	.946
β_{j1}	YEAR	SLOPE	350.080	102	661.98	0.000	.817
β_{j2}	RACE	SLOPE	213.640	97	145.29	0.001	.190
β_{j3}	LUNCH	SLOPE	183.025	96	150.30	0.001	.304

Note. Values in this table were derived from HLM output.

Table 30

Variance Accounted for by Math Achievement Models.

Estimated Parameter Variance

Model	INTERCEPT	LUNCH	RACE
UNCONDITIONAL	1276.619	205.984	246.381
COMPOSITION	919.144 (28.0)	172.730 (16.1)	226.384 (8.1)
THEORETICAL	901.791 (29.4)	183.025 (11.1)	213.640 (13.3)

Note. The number in parentheses represents the percentage of variance accounted for by the HLM model (using the UNCONDITIONAL model as a baseline).

Summary

This chapter reported the results from a test of effective schools theory as described in Chapter 1. The results indicate that the five effective schools characteristics derived from the CSEQ in Chapter 3 had little effect on mean school achievement in reading and math and had little effect on the slopes describing the relationship between important student background characteristics and individual student achievement. The only exceptions to this general pattern were the effect of PARENT INVOLVEMENT on mean reading achievement in schools, the effect of INSTRUCTIONAL OBJECTIVES on the slope relating LUNCH to individual reading achievement within schools, and the effect of PARENT INVOLVEMENT on the slope relating RACE to individual math achievement within schools. Moreover, The effects of effective schools characteristics on within-school slopes, when found, were not entirely consistent with effective schools theory. In one case, the presents of INSTRUCTIONAL OBJECTIVES created more equity among high and low SES students in reading achievement, but this occurred because the focus on instructional objectives tends to lower the reading achievement of higher SES students. In another case, higher levels of PARENT INVOLVEMENT apparently increased inequality in math achievement among students of different racial groups. Neither of these effects would be predicted by effective schools theory. Finally, the analysis demonstrated that the inclusion of the five effective schools characteristics in the HLM models adds little in terms of explaining variance in student achievement.

CHAPTER FIVE: SUMMARY AND CONCLUSIONS

The analyses conducted during the course of this investigation were built upon previous research and procedural improvements. These refinements were motivated in large part by criticisms of earlier school effectiveness research. Thus the present study should be viewed in terms of how it fits into the ever changing agenda of effective schools research. The research questions are also linked in a circular fashion with each other. That is, it was essential that we answered questions about the reliability and validity of the Connecticut School Effectiveness Questionnaire (CSEQ) to lay the foundation for subsequent HLM analyses. But, evidence regarding statistical conclusion validity would not be available until the investigation was complete.

This final chapter opens with a brief discussion of the contributions of the present study to research on school effectiveness. Next, major results from the analyses will be summarized with a focus on the three research questions outlined in Chapter 1. Some implications of these results will follow this summary. Then, in the final section, suggestions for future research in this area will be offered. Though school effectiveness research has evolved a great deal in the last twenty years, it seems obvious that to sustain the progress that has been made, our conceptual and methodological apparatus must also continue to be evaluated and improved.

Summary of Results

The Connecticut School Effectiveness Questionnaire

In view of the continued popularity of the CSEQ as a school assessment instrument, an independent test of its reliability and validity was deemed important enough to raise this issue to the level of a central research question. Therefore, the original seven scales were recreated, using South Carolina data, and their measurement properties thoroughly tested. Before solid inferences can be made regarding the statistical relationship of the school effectiveness measures to achievement, judgements about whether the questionnaire meets the standards for educational research had to be made. The standards being referred to here are those that reflect adequate instrument reliability and validity.

Reliability of the CSEQ scales. Our conclusions regarding the reliability of the CSEQ must be conditioned upon which subscale is being referred to, and the level of analysis used (i.e., teacher or school). At the **teacher level** it was gratifying to find the results of reliability analysis (using Connecticut item-to-construct specifications and the South Carolina data) to be very similar to those provided in the Connecticut Department of Education (1984) handbook. This would suggest that teachers in both states perceive their schools' climate and internal processes in much the same way. That is, both Connecticut and South Carolina teachers seem to share the same notions of what the questionnaire items represent in terms of school characteristics.

The picture changes somewhat when we observe that two of the seven scales (i.e., High Expectations and Time On Task) produce only moderate reliability coefficients as measured by Cronbach's Alpha (ρ_α = .58 and .67 respectively). However, if we consider that Cronbach's Alpha is a <u>minimum</u> estimate of the internal consistency for these scales, this result becomes more acceptable. Also we must keep in mind that we are concerned about the reliability of the instrument at the school level which is quite a different matter from internal consistency.

For the <u>school level</u> reliability analyses, more positive results are generated across all scales. The lowest reliability estimate was .85 (three scale coefficients exceeded .90). With these encouraging results the scales were judged suitable for subsequent analyses. It should be recognized however that scale reliability does not provide a complete picture of the adequacy of the measurement instrument. Though it is relatively easier to establish, reliability does not provide any guarantee that we are measuring the true constructs of interest.

<u>Validity of the CSEQ scales</u>. The validity issue presented a whole new set of questions (and problems) during this investigation. The Connecticut Department of Education (1984) handbook contains information regarding the construct validity of the CSEQ based on multitrait-multimethod (MT-MM) analysis (see Campbell & Fiske, 1959; Marsh, Smith & Barnes, 1983). As the authors of the Connecticut handbook point out "convergent validity refers to a confirmation on the meaning of a trait measured by different methods. The more distinct the methods, the more convergent validity is established" (see Campbell & Fiske, 1959, p. 84). "Discriminant validity refers to the

distinctiveness of the various traits, and it is inferred from the relative lack of correlation among the different traits when compared to the convergence coefficients " (Marsh et al., 1984, p. 335). The question of validity in this sense rests on the degree to which the School Effectiveness Interview (developed by the same Connecticut Department of Education researchers) and CSEQ represent truly distinct methods for measuring the seven effective schools characteristics.

The MT-MM validity analysis was not an option with the South Carolina data (i.e., there was no interview data to match it with). Therefore a principal components analysis was performed and the resulting factors compared with the item to construct specifications recommended by the Connecticut Department of Education. Figure 1 provided an indication that Instructional Leadership (and to a lesser extent Clear School Mission) had better item-to-construct fit than did the remaining scales, with High Expectations having the worst fit between Connecticut item-to-construct recommendations and the actual response categories arrived at through principal components analysis.

Thus, the findings provide only questionable support for the conclusion that the CSEQ reliably measures seven school-level characteristics. While the seven Connecticut scales appear reliable, a principal components analysis found that the South Carolina data did not provide an adequate match between construct and response homogeneity. Moreover, there were high correlations among scales. In fact, some of these scales are so highly correlated (e.g., Frequent Monitoring and Time On Task; High Expectations and Home School Relations) that we questioned whether they truly represent different

constructs from the teachers' point of view. Therefore, the remaining discussion of scale validity will be limited to the five component-based scales chosen for further study in this investigation.

<u>Properties of Final Scales</u>

There are at least three advantages in focusing on our five final scales. First, these scales were reliable at <u>both</u> the teacher and school levels of analysis. Second, these five scales demonstrated the best match between construct homogeneity (as determined by the Connecticut research group) and response homogeneity (from the elementary teachers in Connecticut and South Carolina). Finally, the final five constructs are represented by fewer questionnaire items. This could be of great practical importance as researchers try to find the most efficient means of collecting data from school staffs. The evidence is less clear regarding the construct validity of these five subscales. This matter comes down to how sure we are that the scales measure what their labels and definitions suggest they are measuring. The results here indicate that we can depend on the construct validity of the IL and CSM scales more than the others, but as this answer implies, validity in this sense is a matter of degrees. Evidence for the degree of face and content validity an instrument like the CSEQ possesses is ultimately a matter of intersubjective agreement. More "objective" standards must be satisfied to establish construct and criterion validity.

Support for CSEQ subscale validity in this stronger sense could be provided by the HLM analyses themselves. That is, had we found the predicted relationships between our school characteristics and student

achievement, we could then be assured that the scales measure what they purport to measure. The fact that we do not find these predicted relationships does not, in and of itself, prove the scales invalid (recall the discussion of threats to statistical conclusion validity). One alternative explanation for the apparent absence of effects is that the theory upon which the characteristics are founded is too weak to stand up to the requirements of the HLM analysis. Of course there could also be some combination of questionable validity, specification error, and inadequate theory at work here.

Finally, the research question involving the appropriateness of using the CSEQ as a research tool as well as a needs assessment instrument can tentatively be answered in the affirmative. The strongest support is provided for the validity of the CSM and IL scales, although all of the seven scales are reliable measures of the effective schools characteristics at the school level. More will be said later about improving these properties of the CSEQ in future research efforts.

The "Quality" Issue

The question here is rather straightforward. Do the effective schools characteristics, as measured by the five component- and CSEQ-based scales, have a significant impact on average third grade achievement after we have statistically controlled for student background and school composition variables? The results provided by our analysis seems to suggest an even more straightforward answer: no.

The INTERCEPT for our math outcome measure (which represents the

average achievement of a school in our Math Scale Score model) was not significantly related to any of the school effectiveness characteristics tested. And, only one of the characteristics, INSTRUCTIONAL OBJECTIVES, was related to the INTERCEPT in the Reading Scale Score model. Possible reasons for this "no effect" result will be discussed again in the suggestions for future research section. At this time, however, it seems clear that either the theory upon which the research is based lacks substance, or a widely used assessment instrument (the CSEQ) does not provide valid measurements for the constructs defined here, or both.

The "Equity" Question

An effect on the SES slope was detected when our proxy measure of socioeconomic status (LUNCH) was modeled by the school characteristic INSTRUCTIONAL OBJECTIVES, but this effect was found for only the reading achievement score outcomes. For our reading model, the statistical relationship was such that the SES gap narrows as ratings on the INSTRUCTIONAL OBJECTIVES scale increased in value (all else held constant). But, this closing of the achievement gap is due nearly as much to higher SES students doing less well on the BSAP reading exam, under conditions where a school's rating on the INSTRUCTIONAL OBJECTIVES scale is high, as it is to lower SES students' improved performance under the same conditions (see Figure 6).

With regard to the math scale score model, PARENT INVOLVEMENT has a statistically significant effect on the RACE slope. But, here the relationship is such that the achievement gap between Black students and members of other ethnic groups widens as the school's rating on

PARENT INVOLVEMENT improves (see Figure 10). Again, this is not what our theory predicts for this or any other school effectiveness characteristic.

Contribution of Present Study

Refining the CSEQ

Little information is available regarding the measurement properties of the Connecticut School Effectiveness Questionnaire (CSEQ) beyond that provided by the Connecticut Department of Education. In this study the instrument was evaluated in different surroundings with a sample of over 100 schools and nearly 3,000 elementary teachers. Not only did we confirm the internal consistency estimates derived in the earlier Connecticut research, but for the first time evidence of the scale reliability as measures of these seven characteristics was provided at the school level. Practitioners who use the 1984 version of the CSEQ should be aware that the results of this study and the work done in Connecticut lead us to conclude that the "High Expectation", "Time On Task", and "Frequent Monitoring" scales need to be revised.

During this assessment of the CSEQ, we saw that the scales could be shortened considerably yet still retain the favorable properties of the original instrument. This would indicate that scales could be further strengthened by replacing items that did not survive our selection process with statements that more closely reflect the core constructs of interest, or that an equally reliable and valid instrument could be created that takes only half as long to fill out.

In either case, the "new" instrument should be validated with further research before being used in future school improvement projects.

Improving Analytic Models

The persistent "unit of analysis" problems commonly associated with school effectiveness research have been addressed here with the application of HLM to the problem. This study represents the first time HLM has been used in an analysis that involved as many as five school effectiveness characteristics simultaneously. Thus, two general analytic problems were addressed here. One deals with the use of single-level vs. multi-level regression models. The other issue involves the use of one-variable-at-a-time "shotgun" research vs. a meticulously specified model. The relatively new approach used here (i.e., HLM) has the potential to more accurately model the processes that affect student achievement at multiple organizational levels. At the very least we have reduced errors of interpretation that are introduced as a result of the inherent biases associated with more traditional analytic procedures. HLM is not the answer to every problem that arises in school effectiveness research. No analytic procedure can be. But, as this technique is used more and more in studies like this one, its strengths and weaknesses will be identified. This process should led to even better analytic models.

Implications of the Results

The most important finding of this study may be the lack of effects found. From a researcher's standpoint, one implication of this

result is that a great deal more work is needed in operationalizing the constructs which define internal school processes that actually do affect achievement. This study has shown that it is possible to obtain reliable (if not valid) measures of school characteristics with a paper and pencil assessment instrument. Further, this study has demonstrated once again that measurable differences do exist in the between-schools variance on achievement test scores even after controlling for important student-level variables like SES, ethnic background, sex, and prior achievement.

This study does not provide support for the claims made of the "correlates" of effective schools. Only one of the five effective schools characteristics tested here was significantly related to average achievement. And, only one variable had the anticipated effect of <u>reducing</u> the achievement gap between two identifiable subgroups (i.e., lower and higher SES students). Even then, this "narrowing" effect came at the price of negatively affecting the average performance of the higher SES group. These findings imply that if our intention is to increase achievement in the basic skills areas of reading and math (in typical schools), then focusing on the characteristics as defined by the CSEQ would be inappropriate. This conclusion may only generalize to other elementary schools with similar third grade populations, however.

Suggestions for Future Research

We can learn a great deal from both the strengths of this study and its limitations. It is argued that the HLM approach is the most appropriate at this time for investigating multilevel phenomenon. Researchers who deal with such data (and that means the majority of us) should seriously consider employing this tool in their work, or be able to justify not using it on grounds other than expediency.

HLM, like any regression approach, requires that we make certain assumptions regarding the properties of our data. One possible explanation for the lack of significant statistical relationships reported here involves the skewed achievement distributions associated with the BSAP exams. Though some reduction in the match between curriculum content and what is tested may result, the properties of data derived from standardized, norm-referenced, tests may be better suited to the HLM analysis. Therefore, it is suggested that future studies be conducted with assessment instruments that do not suffer from ceiling effects.

Achievement in the basic skills has been the focus of this and most other studies of this type. The objections of those who warn us about the dangers of too narrow a focus should be heeded. Important work still needs to be done within the basic skills arena, but this should not prevent us from including other outcome measures, especially if the content of these measures is well matched to the curriculum of the schools under investigation.

There is a need to measure variables at the ability group,

classroom, and even district level along with other school-level variables to better account for the effects on student achievement from all these sources and reduce errors of estimation. It is vital that the research design allows the investigator to link students to these various organizational levels in order to sort out the impact that each might have on their achievement. The restriction of not matching students with their teachers in this study seriously impaired our ability to explore possible explanations for the observed results. Some examples of this problem will be given later in this section. In the same vein, the design of this study could have been strengthened considerably had we been able measure the predictor and criterion variables at the same point in time, and over a series of time intervals.

Further, there may be other school characteristics that are far more important in terms of achievement outcomes that were overlooked. HLM regression analyses indicate that less than 5% of the total variation in the outcome was accounted for by our effective schools characteristics after controlling for student background variables (see Tables 24 and 30). For reading achievement, our "theoretical" model leaves over 40% of the between-school parameter variance unaccounted for. The situation is even more pronounced with regard to math achievement. Here our final model left over 70% of the variance in math scale scores unaccounted for. Must we assume that the remaining variance represents random error?

In light of our findings regarding the INSTRUCTIONAL OBJECTIVES and PARENT INVOLVEMENT constructs it might be helpful to obtain the

students' and parents' perspectives on these constructs. This approach has been used in earlier effective schools research and should be refined and continued. Recognizing the challenges this may present (especially at the primary grade levels), researchers may find that the benefits of such efforts will far outweigh the costs. In this study, for example, parent-supplied data could have provided answers about differential application of the PARENT INVOLVEMENT characteristic to particular subgroups of students. Student reports could help us explore other areas. For example, do teachers with a preponderance of low SES students in their classes view Positive School Climate more as a question of safety and discipline than a "positive learning environment" issue? Obtaining student and parent responses and linking them with their teachers' profile would seem the most direct way to address such questions.

Finally, reducing errors of interpretation regarding what the CSEQ actually measures seems central to generating valid conclusions from our research. Teachers who take part in school improvement projects by responding to questionnaires should be well versed in terms of the constructs involved. This in turn should improve both the reliability and validity of the results. At the school level the respondent, and not the survey he or she responds to, becomes the measurement instrument. The problem identified here reduces to an matter of quantitative vs. qualitative evaluation. The analyses and instrument used here clearly put this study on the quantitative side of this issue. The question may not be whether "most parents would rate this school as superior", but why. It may not be a matter of the number of

classroom observations made each year, but what is done with the information collected during this process. Whether the principal "requires and regularly reviews lesson plans" may not be as important as how useful that review is in terms of boosting student achievement scores. So with respect to improving the criterion validity of the CSEQ these more qualitative questions must be answered, if possible.

A Final Note

The suggestions for future studies outlined here may seem to assume that researchers have unlimited resources to work with. But, the issue is not how to raise huge sums of money to support elaborate research projects. The challenge lies in finding ways to incorporate what we know to be the best procedures into existing research and school improvement efforts. The bottom line is that research on schools and school improvement projects could <u>both</u> benefit greatly by working much more closely with each other. Generating guidelines for "what works" in schools is the easy part. Faithfully implementing them and evaluating their true impact is where we should be focusing our efforts.

REFERENCES

References

Achen, C. H. (1982). Interpreting and using regression. Sage University Papers on Quantitative Applications in the Social Sciences, 07-029. Beverly Hills, CA: Sage Publications.

Anderson, L. W. & Mandeville, G. K. (1986, April). Toward the solution of problems inherent in the identification of effective schools on a statewide basis. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA.

Austin, G. R. (1981). Exemplary schools and their identification. New Directions for Testing and Measurement, 10, 31-48.

Berliner, D. C. (1979). Tempus educare. In P. L. Peterson & H. J. Walberg (Eds.), Research on teaching: Concepts, findings and implications. Berkeley, CA: McCutchan.

Berliner, D. C. & Fisher, C. W. (1985). One more time. (pp. 333-347) In C. W. Fisher and D. C. Berliner (Eds.) Perspectives on instructional time. White Plains, NY: Longman Inc.

Berry, W. D. & Feldman, S. (1985). Multiple Regression in practice. Beverly Hills, CA: Sage Press #50.

Block, A. W. (1983). Effective schools: A summary of research. Research Brief. Arlington, VA: Educational Research Service Inc.

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

Bowles, S. S. & Levin, H. M. (1968). The determinants of scholastic achievement: An appraisal of some recent evidence. Journal of Human Resources, 3(1), 2-24.

Brookover, W. B., Beady, C. H., Flood, P. K., Schweitzer, J. H., & Wisenbaker, J. M. (1979). School social systems and student achievement: Schools can make a difference. New York, NY: Praeger.

Brookover, W. B. & Lezotte, L. W. (1977). Changes in school characteristics coincident with changes in student achievement (Executive summary). East Lansing, MI: College of Urban Development, Michigan State University.

Brookover, W. B. & Schneider, J. M. (1975). Academic environments and elementary school achievement. *Journal of Research and Development in Education, 9*(1), 82-91.

Brookover, W. B., Schweitzer, J. H., Schneider, J. M., Beady, C. H., Flood, P. K. & Wisenbaker, J. M. (1978). Elementary school social climate and school achievement. *American Educational Research Journal, 15*(2), 301-318.

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631-661.

Brophy, J. E. & Good, T. L. (1986). *Teacher behavior and student achievement.* (pp. 328-375) In M. C. Wittrock (Ed.) Handbook of Research on Teaching (Third Edition). New York, NY: Macmillan Publishing Company.

Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. T. (1986). An Introduction to HLM: Computer program and users' guide (Version 1.0).

Burstein, L. (1980). *The role of levels of analysis in the specification of educational effects.* (pp. 119-190) in R. Dreeben and J. A. Thomas (Eds.) The Analysis of Educational Productivity, Vol. I: Issues in Microanalysis. Cambridge, MA: Ballinger.

Burstein, L. (1987, April). Speech entitled: *The measurement of effective schooling - Some nagging issues.* Presented at the Annual Meeting of the American Educational Research Association, before the School Effectiveness SIG. Washington, DC.

Burstein, L. & Miller, M. D. (1980). Regression-based analysis of multi-level educational data. *New Directions for Methodology of Social and Behavioral Sciences, 6,* 194-211.

California State Department of Education (1977). California school effectiveness study, The first year: 1974-75 A report to the California legislature as required by education code section 12851. Office of Program Evaluation and Research.

Carroll, J. (1963). A model for school learning. *Teacher College Record, 64,* 723-733.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Clark, T. A. & MacCarthy, D. P. (1983). School improvement in New York city: The evolution of a project. Educational Researcher, 12(4), 17-24.

Cohen, M. (1982). Effective schools: Accumulating research findings. American Education, Jan-Feb. 13-16.

Cohen, S. A. (1987). Instructional alignment: Searching for a magic bullet. Educational Researcher, 16(8), 16-20.

Coleman, J. S., Campbel, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Winfield, F. D. & York, R. L. (1966). Equality of educational opportunity. Washington, DC: U. S. Department of Health Education and Welfare.

Coleman, J. S., Hoffer, T. & Kilgore S. B. (1982). High school achievement: Public, Catholic, and private schools compared. New York, NY: Basic Books.

Connecticut Department of Education (1984). Handbook for use of the Connecticut school effectiveness interview and questionnaire. Hartford, CT.

Consumer Information Center (May, 1986). Research findings you can trust. Instructor, 46-47.

Cooley, W. W., Bond, L. & Mao, B. (1981). Analyzing multi-level data, in R. A. Berk, (Ed.). Educational Evaluation Methodology, Baltimore, MD: Johns Hopkins University Press.

Crano, W. D. & Mellon, P. M. (1978). Causal influence of teachers' expectations on children's academic performance: A cross-lagged panel analysis. Journal of Educational Psychology, 70(1), 39-49.

Cronbach, L. J. (1976). Research on classrooms and schools: Formulation of questions, design, and analysis. Occasional paper of the Stanford Evaluation Consortium.

Cuban, L. (1983). Effective schools: A friendly but cautionary note. Phi Delta Kappan, 64(10), 695-696.

Dayton & Schafer (1973). [see Kirk, 1982].

Denham, C. & Lieberman, A. (1980). Time to learn. Washington, DC: National Institute of Education.

Dyer, H. S. (1966). The Pennsylvania plan. Science Foundations, 50, 206-211.

Dyer, H. S. (1970). Toward objective criteria of professional accountability in the schools of New York city. Phi Delta Kappan, 52, 206-211.

Dyer, H. S., Linn, R. L. & Patton, M. J. (1969). A comparison of four methods for obtaining discrepancy scores based on observed and predicted school system means on achievement tests. American Educational Research Journal, 6, 591-605.

Edmonds, R. R. (1981). Invitational address to the general session of the Annual Meeting of the Association for Supervision and Curriculum Development. St. Louis, MO: ASCD.

Edmonds, R. R. (1982). Programs for school improvement: An overview. Educational Leadership, 40(3), 4-11.

Firestone, W. A. & Herriott, R. E. (1982). Perscriptions for effective elementary schools don't fit secondary schools. Educational Leadership, 40(3), 51-53.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R. S., Cahen, L. S. & Dishaw, M. M. (1980). Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington, DC: Department of Education, NIE.

Floden, R., Porter, A., Schmidt, W., Freeman, D., & Schwille, J. (1981). Response to curriculum pressures: A policy capturing study of teacher decisions about content. Journal of Educational Psychology, 73(2), 129-141.

Frechling, J. A. (1982, March). Alternative methods for determining effectiveness: Convergence and divergence. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

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

Gauthier, W. J., Pecheone, R. L. & Shoemaker, J. (1985). Schools can be more effective. Journal of Negro Education, 54(3), 388-408.

Glasman, N. S. (1984). Student achievement and the school principal. Educational Evaluation and Policy Analysis, 6(3), 283-296.

Good, T. & Brophy, J. E. (1986). School effects. (pp. 570-602) In M. C. Wittrock (Ed.) Handbook of Research on Teaching (Third Edition). New York, NY: Macmillan Publishing Company.

Gottfredson, D. C., Hybl, L., Gottfredson, G. & Castaneda, R. (1986, April). School climate assessment instruments: A review. Paper presented at the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 278 702).

Huynh, H. & Casteel, J. (1983, March). Technical works for Basic Skills Assessment Programs. University of South Carolina.

Iverson, B. K., Brownlee, G. D., & Walberg, H. J. (1981). Parent -teacher contacts and student learning. Journal of Educational Research, 74(6), 394-396.

Jencks, C. S., Smith, M., Ackland, H., Bane, M. J., Cohen, D., Gintis, H., Heynes, B. & Michelson, S. (Eds.). (1972). Inequality: A reassessment of the effects of family and schooling in America. New York, NY: Basic Books.

Karweit, N. L. (1983). Time on task: A research review. Baltimore, MD: Center for Social Organization of Schools, The Johns Hopkins University.

Karweit, N. L. (1985). Should we lengthen the school term? Educational Researcher, 14(6), 9-15.

Kim, J. O. & Mueller, C. W. (1978). Introduction to factor analysis. Beverly Hills, CA: Sage Press.

Kirk, R. E. (1982). Experimental design: Procedures for the behavioral sciences (Second Edition). Belmont, CA: Brooks/Cole Publishing Company.

Klitgaard, R. E. & Hall, G. R. (1973). A statistical search for unusually effective schools. Santa Monica, CA: RAND.

Knapp, T. R. (1977). The unit of analysis problem in applications of simple correlational analysis to educational research. Journal of Educational Statistics, 2(3), 171-186.

Lee, V. E. (1986, April). Multi-level causal models for social class and achievement. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Lee, V. E. & Bryk, A. S. (1986, April). The effects of high school academic organization on the social distribution of achievement. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Lezotte, L. W. (1986, April). School effectiveness: Reflections and future directions. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Lezotte, L. W., Hathaway, D. V., Miller, S. K., Passalacqua, J. & Brookover, W. B. (1980). School learning climate and student achievement: A social systems approach to increased student learning. Teacher Corps, Florida State University.

Mackenzie, D. E. (1983). Research for School Improvement: An appraisal of some recent trends. Educational Researcher, 12(4), 5-17.

Madaus, G. F., Airasian, P. W., & Kellaghan, T. (1980). School effectiveness: A reassessment of the evidence. New York, NY: McGraw-Hill Book Company.

Madaus, G. F., Kellaghan, T.,Rakow, E. A. & King, D. J. (1979). The sensitivity of measures of school effectiveness. Harvard Educational Review, 49(2), 207-230.

Marsh, H., Smith, I. & Barnes, J. (1983). Multitraite-multimethod analyses of the self-description questionnaire: Student-teacher agreement on multidimensional ratings of student self concept. American Education Research Journal, 20, 333-357.

Maryland State Department of Education (1978, February). Process evaluation: A comprehensive study of outliers. Baltimore, MD: Center for Educational Research and Development, University of Maryland. (ERIC Document Reproduction Service No. ED 160 644).

Marzano, R. J., Guzzetti, B. J. & Hutchins, C. L. (1984). A study of selected school effectiveness variables: Some correlates that are not causes. (ERIC Document Reproduction Service No. ED 253 328).

McCormack-Larkin, M. & Kritek, W. J. (1983). Milwaukee's Project RISE. Educational Leadership, 40(3), 16-21.

Meyer, W. J. (1985). Summary, Integration, and prospective. In J. B. Dusek, V. C., Hall, & W. J. Meyer (Eds.), Teacher expectancies. Hillsdale, NJ: Lawrence Erlbaum.

Miles, M. B. & Kaufman, T. (1985). A directory of programs. In R. M. J. Kyle (Ed.), Reaching for excellence: An effective schools sourcebook. Washington, DC: U. S. Government Printing Office.

Miller, S. K. (1985). Research on exemplary schools: An historical perspective. In G. R. Austin & H. Garber (Eds.), Research on exemplary schools (pp. 3-30). Orlando, FL: Academic Press.

Mitman, A. L. & Snow, R. E. (1985). Logical and methodological problems in teacher expectancy research. In J. B. Dusek, V. C., Hall, & W. J. Meyer (Eds.), Teacher expectancies. Hillsdale, NJ: Lawrence Erlbaum.

Mosteller, F. & Moynihan, D. P. (1972). On equality of educational opportunity. Papers deriving from the Harvard University Faculty Seminar on the Coleman Report. New York, NY: Vintage Books.

Murphy, J. (1988). Methodological, measurement, and conceptual problems in the study of instructional leadership. Educational Evaluation and Policy Analysis, 10(2), 117-139.

Natriello, G. & Dornbush, S. M. (1984). Teacher evaluative standards and student effort. New York, NY: Longman.

Niedermeyer, F. C. & Yelon, S. (1981). Los Angeles aligns instruction with essential skills. Educational Leadership, 38, 618-620.

Odden, A. & Dougherty, V. (1982). State programs of school improvement: A 50-state survey. Denver, CO: Education Commission of the States, Educational Programs Division. (ERIC Document Reproduction Service No. ED 224 135).

Ogawa, R. T. & Hart, A. W. (1985). The effect of principals on the instructional performance of schools. The Journal of Educational Administration, 23(1), 59-72.

Porter, A. C. (1983). The role of testing in effective schools: How can testing improve school effectiveness? American Education, Jan-Feb, 25-28.

Porter, A. C. & Raudenbush, S. W. (1987). Analysis of covariance: its model and use in psychological research. Journal of Counseling Psychology, 34(4), 383-392.

Purkey, S. C. & Smith, M. S. (1983). Effective schools: A review. The Elementary School Journal, 83(4), 427-452.

Purkey, S. C. & Smith, M. S. (1985). School Reform: The district policy implications of the effective schools literature. The Elementary School Journal, 85(3), 353-389.

Ramey, M. (1987, April). Causal models relating school characteristics to a measure of school contributions to academic achievement. Paper presented at the Annual Meeting of the American Educational Research Association. Washington, DC.

Ramey, M. & Hillman, L. (1983). School characteristics related to student academic growth. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 230 601).

Raudenbush, S. W. & Bryk, A. S. (in press). Methodological advances in studying effects of schools and classrooms on student learning. To appear in Review of Research in Education.

Raudenbush, S. W. & Bryk, A. S. (1986). A hierarchical model for studying school effects. _Sociology of Education_, _59_, 1-17.

Richards, D. M. (1986, April). Productive and effective schools. Paper presented at the Annual Meeting of the American Education Finance Association. Chicago, IL.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. _American Sociological Review_, _15_, 351-357.

Rogosa, D. (1980). A critique of cross-lagged correlations. _Psychological Bulletin_, _88_, 245-258.

Rosenshine, B. V. & Berliner, D. C. (1978). Academic engaged time. _British Journal of Teacher Education_, _4_(1), 3-16.

Rosenthal, R. (1987). Pygmalion effects: Existence, magnitude, and social importance (a reply to Wineburg). _Educational Researcher_, _16_(9), 37-41.

Rosenthal, R. & Jacobson, L. (1968). _Pygmalion in the classroom_. New York, NY: Holt, Rinehart & Winston.

Rossmiller, R. A. (1982, September). _Managing school resources to improve student achievement_. Paper presented at the State Superintendent Conference for District Administrators, Madison, WI.

Rowan, B. (1985). _The assessment of school effectiveness_. In R. M. J. Kyle (Ed.) Reaching for excellence: An effective schools sourcebook. Washington, DC: US Government Printing Office.

Rowan, B., Bossert, S. T. & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. _Educational Researcher_, _12_(4), 24-31.

Shoemaker, J. (1984). Reseach-based school improvement practices. Hartford, CT: Connecticut Department of Education.

Sirois, H. A. & Villinova, R. M. (1982, March). _Theory into practice: A theoretical and research base for the characteristics of effective schools_. Paper presented at the Annual Meeting of the American Educational Research Association. New York, NY. (ERIC Document Reproduction Service No. ED 217 558).

Slavin, R. E. (1980). Effects of individual learning expectations on student achievement. _Journal of Educational Psychology_, _72_(4), 520-524.

Stedman, L. C. (1985). A new look at the effective schools literature. _Urban Education_, _20_(3), 295-326.

Stringfield, S. & Teddlie, C. (1987, April). A time to summarize: Six years and three phases of the Louisiana school effectiveness study. Paper presented at the Annual Meeting of the American Educational Research Association. Washington, DC.

Sweeney, J. (1982). Research synthesis on effective school leadership. Educational Leadership, 39, 346-352.

United States Department of Education (1986). What works: Research about teaching and learning. Washington, DC.

Weber, G. (1971). Inner-city children can be taught to read: Four successful schools. Washington, DC: Council for Basic Education.

Weinberg, S. S. (1987). The self-fulfillment of the self-fulfilling prophecy: A critical appraisal. Educational Researcher, 16(9), 28-37.

Zirkel, P. A. & Greenwood, S. C. (1987). Effective schools and effective principals: Effective research? Teachers College Record, 89(2), 255-267.

APPENDIX A

CODE BOOK FOR SOUTH CAROLINA DATA SET

CODE BOOK

"Between-Schools" Variables

SID   Arbitrary three digit number (range 101 to 219) representing elementary schools with third grade in their organizational structure in South Carolina.

%LUNCH   Percentage of students in each school on the free or reduced price lunch program. Mean - 52.8 Range 5.9 to 97.8

%BLACK   Percentage of black students in each school. Mean - 43.2 Range 2.2 to 99.7

%MALE   Percentage of males in each school in the sample. Mean - 51.4 Range 43.1 to 62.0

%REPEAT   Percentage of students who repeated a grade. Scores on the achievement measure used as the outcome variable in this study have no bearing on the "repeater" classification. Mean -6.5 Range - .2 to 21.8

%GIFTED   Percentage of students in a particular school classified as gifted or talented. Again, this classification is not related in any way with the outcome measure. The criterion used to so classify students dose differ from one school district to the next, however. Mean - 4.9 Range 0 to 33.3

SIZE   The number of teachers from each school in the sample who responded to the Connecticut school effectiveness questionnaire. Mean - 28.02 Range 9 to 54

ORDERLY ENVIRONMENT
     Mean score for each school on the overlapping items of the "Safe and Orderly Environment" scale and component C-1 (see Figure 2).

INSTUUCTIONAL OBJECTIVES
     Mean value for each school on the overlapping items of the "Clear School Mission" scale and component C-2 (Figure 2).

PRINCIPAL LEADERSHIP
     Mean score for the overlapping items of the "Instructional Leadership" scale and component C-3 (Figure 2).

MONITORING & FEEDBACK
     Mean score on the combined "Opportunity to Learn / Time On Task" and "Frequent Monitoring" scale intersection with component C-5 (Figure 2).

PARENT INVOLVEMENT
>Mean score for overlapping items of the "Home School
Relations" scale and component C-4 (Figure 2).

"Within-Schools" Variables

LUNCH   Individual measure of participation in the school's free or
            reduced price lunch program (coded lower SES — -.507, higher
            SES — .493).

RACE   Blacks are coded -.575, others — .425.

SEX   Males are coded -.493, females — .507.

REPEAT   Students held back are coded -.933, non-repeaters — .067.

GIFTED   Students classified by their school district as "Gifted or
            Talented" are coded .926, not gifted — -.074.

RSS   Basic Skills Assessment Program (BSAP) reading scale score
            (700 is the standard for passing).

MSS   BSAP Math scale score (700 is the standard for passing).

YEAR   Linear contrast coefficients for year in which BSAP data was
            collected (1984 — -1, 1985 — 0, and 1986 — 1).

APPENDIX B

CONNECTICUT SCHOOL EFFECTIVENESS QUESTIONNAIRE (CSEQ)

ELEMENTARY/MIDDLE SCHOOL
Faculty/Staff Attitudes Toward School Effectiveness Questionnaire

INTRODUCTION

This Questionnaire is one component of the School Effectiveness Assessment Process. Items are drawn from research on school and instructional effectiveness. The school effectiveness characteristics assessed through this Questionnaire are the focal point of the School Effectiveness Process.

The purpose of this Questionnaire is to survey your perceptions based on your own experience with this school. There are no right or wrong answers. Whenever possible, questions are designed to measure "school effects" and you will be asked to generalize about the people working in this school. Where this is not possible, you should respond from your own experience.

Responses are summarized and will be reported back to the faculty, administration, parents, and students of this school in group profile form. To ensure confidentiality, do not write your name on the answer sheet.

INSTRUCTIONS

1. All questions have five (5) possible responses. Record your answer by circling the appropriate number, or if provided, by marking the appropriate number on the computer sheet. The response categories for each item are:

    1 = (SD) Strongly Disagree
    2 = (D) Disagree
    3 = (U) Uncertain, Undecided - This response should be
        used as infrequently as possible.
    4 = (A) Agree
    5 = (SA) Strongly Agree

2. Although some questions may seem to warrant a Yes-No response, the response categories permit you to indicate the intensity of your feelings in relation to the item.

3. Your perceptions based on your own experience with this school are important.

4. Tour interpretation of each item is important. Try to answer the question in respect to the school as a whole, if possible.

5. Each item, must be read carefully. There is not a time limit. Completion of this Questionnaire is expected to take approximately fourty-five (45) minutes.

ELEMENTARY/MIDDLE SCHOOL
Faculty/Staff Attitudes Toward School Effectiveness Questionnaire

DIRECTIONS: Read over each of these answer categories carefully. Then answer each of the questions by circling the proper answer or, if provided, by marking the computer answer sheet.

		SD	D	U	A	SA
1.	This school is a safe and secure place to work.	1	2	3	4	5
2.	In reading, written sequential objectives exist up through all grades.	1	2	3	4	5
3.	In this school, low-achieving students present more discipline problems than other students.	1	2	3	4	5
4.	Most problems facing this school can be solved by the principal and faculty without a great deal of outside help.	1	2	3	4	5
5.	Most students in this school are eager and enthusiastic about learning.	1	2	3	4	5
6.	The principal makes several formal classroom observations each year.	1	2	3	4	5
7.	Discussions with the principal often result in some aspect of improved instructional practice.	1	2	3	4	5
8.	The physical condition of this school is generally unpleasant and unkempt.	1	2	3	4	5
9.	Most parents would rate this school as superior.	1	2	3	4	5
10.	The principal reviews and interprets test results with and for the faculty.	1	2	3	4	5
11.	School-wide objectives are the focal point of reading instruction in this school.	1	2	3	4	5
12.	In reading, initial skill instruction is often presented to a heterogeneous group of students.	1	2	3	4	5
13.	Instructional issues are seldom the focus of faculty meetings.	1	2	3	4	5
14.	Pull out programs (e.g. Chapter I, Special Ed., Gifted, etc.) often disrupt and interfere with basic/essential skills instruction.	1	2	3	4	5

		SD	D	U	A	SA
15.	Mathematics objectives are not coordinated and monitored up through all grades in this school.	1	2	3	4	5
16.	The principal uses test results to recommend modifications or changes in the instructional program.	1	2	3	4	5
17.	There is clear, strong, centralized instructional leadership from the principal in this school.	1	2	3	4	5
18.	Ninety-five to one hundred percent of the students in this school can be expected to complete high school.	1	2	3	4	5
19.	The principal regularly gives feedback to teachers concerning lesson plans.	1	2	3	4	5
20.	Criterion-referenced tests are not used to assess basic skills throughout the school.	1	2	3	4	5
21.	At the principal's initiatives, teachers work together to effectively coordinate the instructional program within and between grades.	1	2	3	4	5
22.	In basic/essential skills instruction in this school, reteaching and specific skills remediation are important parts of the teaching process.	1	2	3	4	5
23.	A positive feeling permeates the school.	1	2	3	4	5
24.	All materials and supplies necessary for instruction in basic/essential skills are available.	1	2	3	4	5
25.	Staff and students do not view security as a problem in this school.	1	2	3	4	5
26.	In reading, an identified set of objectives or skills that all students must master exists at each grade level.	1	2	3	4	5
27.	Teachers in this school do not hold consistently high expectations of all students.	1	2	3	4	5
28.	When students are assigned seat work, teachers monitor it closely.	1	2	3	4	5
29.	In mathematics, there is an identified set of objectives or skills that all students must master at each grade level.	1	2	3	4	5
30.	There are few student interruptions during class time.	1	2	3	4	5

		SD	D	U	A	SA
31.	Homework is monitored at home and in school with follow-up.	1	2	3	4	5
32.	A written statement of purpose that is the driving force behind most important decisions does not exist in this school.	1	2	3	4	5
33.	Outside interruptions do not often interfere with basic/essential skill instruction in this school.	1	2	3	4	5
34.	During parent-teacher conferences, there is a focus on factors directly related to student achievement and basic/essential skills mastery.	1	2	3	4	5
35.	Almost all students complete assigned homework before coming to school.	1	2	3	4	5
36.	Ninety to one hundred percent of your students parents attend scheduled parent-teacher conferences.	1	2	3	4	5
37.	The standardized testing program is an accurate and valid measure of the basic skills curriculum in this school.	1	2	3	4	5
38.	Almost all students are expected to master basic/ essential skills at each grade level.	1	2	3	4	5
39.	In Language Arts, an identified set of objectives or skills that all students must master exist at each grade level.	1	2	3	4	5
40.	Teachers and the principal thoroughly review and analyze test results to plan instructional program modifications.	1	2	3	4	5
41.	In Language Arts, school-wide objectives are the focus of instruction in this school.	1	2	3	4	5
42.	Standardized test results are not used to evaluate program objectives.	1	2	3	4	5
43.	Multiple assessment methods are used to assess student progress in basic/essential skills (e.g., criterion-referenced tests, work samples, mastery check lists, etc.).	1	2	3	4	5
44.	Teachers, administrators and parents assume responsibility for discipline in this school.	1	2	3	4	5
45.	The principal requires and regularly reviews lesson plans.	1	2	3	4	5

	SD	D	U	A	SA

46. The principal is very active in securing resources, arranging opportunities and promoting staff development activities for the faculty. 1 2 3 4 5

47. There is little cooperation in regard to homework monitoring between parents and teachers in this school. 1 2 3 4 5

48. The principal leads frequent formal discussions concerning instruction and student achievement. 1 2 3 4 5

49. Parent-teacher conferences seldom result in specific plans for home-school cooperation aimed at improving student classroom achievement. 1 2 3 4 5

50. Reading objectives are coordinated and monitored through all grades. 1 2 3 4 5

51. In mathematics, instruction is often presented to homogeneous ability groups. 1 2 3 4 5

52. Two hours or more are allocated for reading/language arts each day throughout this school. 1 2 3 4 5

53. Specific feedback on daily assignments is given regularly and followed by the teacher. 1 2 3 4 5

54. In Language Arts, written, sequential objectives do exist at each grade level. 1 2 3 4 5

55. Individual teachers determine allocated time for basic/ essential skill instruction without guidelines or discussion from the administration. 1 2 3 4 5

56. The school building in neat, bright, clean and comfortable. 1 2 3 4 5

57. Formal observations by the principal are regularly followed by a post observation conference. 1 2 3 4 5

58. There is no systematic, regular assessment of student's basic/essential skills in most classrooms. 1 2 3 4 5

59. Language Arts objectives are coordinated and monitored through all grades. 1 2 3 4 5

60. The principal frequently communicates to individual teachers their responsibility in relation to student achievement. 1 2 3 4 5

	SD	D	U	A	SA

61. Student assessment information (such as criterion referenced tests, skills checklists, etc.) is regularly used to give specific feedback and plan appropriate instruction.
    1   2   3   4   5

62. Special instructional programs for individual students are integrated with classroom instruction and the curriculum.
    1   2   3   4   5

63. The principal does not put much emphasis on the meaning and use of standardized test results.
    1   2   3   4   5

64. Very few parents of students in your class visit the school to observe the instructional program.
    1   2   3   4   5

65. Typical daily lessons in this school follow this sequence: teacher presentation, student practice, specific feedback, evaluation of student performance.
    1   2   3   4   5

66. The principal regularly brings instructional issues to the faculty for discussion.
    1   2   3   4   5

67. Most parents understand and promote the school's instructional program.
    1   2   3   4   5

68. There is active parent organization in this school that involves many parent.
    1   2   3   4   5

69. Many parents are involved in an overall home/school support network.
    1   2   3   4   5

70. Supervision is directed at instruction.
    1   2   3   4   5

71. Teachers and parents are aware of the homework policy in this school.
    1   2   3   4   5

72. Fifty minutes or more is allocated to mathematics instruction each day in this school.
    1   2   3   4   5

73. Student assignments in basic/essential skill areas are corrected daily.
    1   2   3   4   5

74. Teachers believe that all students in this school can master basic/essential skills as a direct result of the instructional program.
    1   2   3   4   5

75. The principal is highly visible throughout the school.
    1   2   3   4   5

76. Teachers in this school expect and plan assignments so that students will be highly successful during practice work following direct instruction.
    1   2   3   4   5

	SD	D	U	A	SA
77. Teachers in this school believe they are responsible for all students mastering all basic/essential skills at each grade level.	1	2	3	4	5
78. The principal is accessible to discuss matters dealing with instruction.	1	2	3	4	5
79. Many parents initiate many contacts with this school.	1	2	3	4	5
80. Low-achieveing students answer questions as often as other students in this school.	1	2	3	4	5
81. Student behavior is generally positive in this school.	1	2	3	4	5
82. The principal is an important instructional resource person in this school.	1	2	3	4	5
83. The number of low-income children retained in grade is proportionately equivalent to other children in grade.	1	2	3	4	5
84. Phone calls, newsletters, regular notes, and parent-teacher conferences are ways that most teachers communicate with parents in this school.	1	2	3	4	5
85. Teachers believe that a student's home background is not the primary factor that determines individual student achievement.	1	2	3	4	5
86. During follow-up to formal classroom observations, a plan for improvement frequently results.	1	2	3	4	5
87. Students in this school abide by school rules.	1	2	3	4	5
88. Teachers in this school do not turn to the principal with instructional concerns or problems.	1	2	3	4	5
89. Class atmosphere in this school is generally very conducive to learning for all students.	1	2	3	4	5
90. During follow-up to formal observations, the principal's main emphasis is on instructional improvement.	1	2	3	4	5
91. In mathematics, school-wide objectives are the focal point of instruction.	1	2	3	4	5
92. In this school there is annual standardized testing at each grade level.	1	2	3	4	5
93. The principal rarely makes informal contacts with students and teachers around the school.	1	2	3	4	5

	SD	D	U	A	SA

94. Criterion-referenced tests are used to give specific student feedback in basic/essential skills areas throughout the school.

    1 2 3 4 5

95. Generally, discipline is not a problem in this school.

    1 2 3 4 5

96. Beyond parent conferences and report cards, teachers in this school have several other ways for communicating student progress to parents.

    1 2 3 4 5

97. Individual teachers and the principal do not meet regularly to discuss what the principal will observe during a classroom observation.

    1 2 3 4 5

98. During basic/essential skills instruction, students are working independently on seat work for the majority of allocated time.

    1 2 3 4 5

99. In mathematics, written sequential objectives exist in all grades.

    1 2 3 4 5

100. Students not mastering basic/essential skills are frequently retained in grade.

    1 2 3 4 5

APPENDIX C

CONNECTICUT DEPARTMENT OF EDUCATION (1984)
SCHOOL EFFECTIVENESS SCALES

Table C-1

Positive School Climate (PSC) Items.

1. This school is a safe and secure place to work

5. Most students in this school are eager and enthusiastic about learning

8. The physical condition of this school is generally unpleasant and unkempt

23. A positive feeling permeates the school

25. Staff and students do not view security as a problem in this school

44. Teachers, administrators and parents assume responsibility for discipline in this school

56. The school building in neat, bright, clean and comfortable

81. Student behavior is generally positive in this school

87. Students in this school abide by school rules

95. Generally, discipline is not a problem in this school

_____

Note. The coding for item number 8 was reversed to match the rest of the scale prior to reliability checks.

Table C-2

Clear School Mission (CSM) Items.

2. In reading, written sequential objectives exist up through all grades

11. School-wide objectives are the focal point of reading instruction in this school

15. Mathematics objectives are not coordinated and monitored up through all grades in this school

24. All materials and supplies necessary for instruction in basic/essential skills are available

26. In reading, an identified set of objectives or skills that all students must master exists at each grade level

29. In mathematics, there is an identified set of objectives or skills that all students must master at each grade level

32. A written statement of purpose that is the driving force behind most important decisions does not exist in this school

39. In Language Arts, an identified set of objectives or skills that all students must master exist at each grade level

41. In Language Arts, school-wide objectives are the focus of instruction in this school

50. Reading objectives are coordinated and monitored through all grades

54. In Language Arts, written, sequential objectives do exist at each grade level

59. Language Arts objectives are coordinated and monitored through all grades

91. In mathematics, school-wide objectives are the focal point of instruction

99. In mathematics, written sequential objectives exist in all grades

---

Note. The coding for items 15, 32, and 54 was reversed to match the rest of the scale.

Table C-3

Strong Instructional Leadership (IL) Items.

4. Most problems facing this school can be solved by the principal
    and faculty without a great deal of outside help

6. The principal makes several formal classroom observations each
    year

7. Discussions with the principal often result in some aspect of
    improved instructional practice

10. The principal reviews and interprets test results with and for the
    faculty

13. Instructional issues are seldom the focus of faculty meetings

16. The principal uses test results to recommend modifications or
    changes in the instructional program

17. There is clear, strong, centralized instructional leadership from
    the principal in this school

19. The principal regularly gives feedback to teachers concerning
    lesson plans

21. At the principal's initiative, teachers work together to
    effectively coordinate the instructional program within and
    between grades.

45. The principal requires and regularly reviews lesson plans

46. The principal is very active in securing resources, arranging
    opportunities and promoting staff development activities for the
    faculty

48. The principal leads frequent formal discussions concerning
    instruction and student achievement

57. Formal observations by the principal are regularly followed by a
    post observation conference

60. The principal frequently communicates to individual teachers their
    responsibility in relation to student achievement

66. The principal regularly brings instructional issues to the faculty
    for discussion

70. Supervision is directed at instruction

75. The principal is highly visible throughout the school

Table C-3 (cont'd)

78. The principal is accessible to discuss matters dealing with instruction

82. The principal is an important instructional resource person in this school

86. During follow-up to formal classroom observations, a plan for improvement frequently results

88. Teachers in this school do not turn to the principal with instructional concerns or problems

90. During follow-up to formal observations, the principal's main emphasis is on instructional improvement

93. The principal rarely makes informal contacts with students and teachers around the school

97. Individual teachers and the principal do not meet regularly to discuss what the principal will observe during a classroom observation

---

Note. Responses to items 13, 63, 88, 93, and 97 were recoded to match the polarity of other items in this scale.

Table C-4

High Expectations (HE) Items.

3. In this school, low-achieving students present more discipline problems than other students

12. In reading, initial skill instruction is often presented to a heterogeneous group of students

18. Ninety five to one hundred percent of the students in this school can be expected to complete high school

27. Teachers in this school do not hold consistently high expectations of all students

38. Almost all students are expected to master basic/essential skills at each grade level

51. In mathematics, instruction is often presented to homogeneous ability groups

74. Teachers believe that all students in this school can master basic/essential skills as a direct result of the instructional program

77. Teachers in this school believe they are responsible for all students mastering all basic/essential skills at each grade level

80. Low-achieveing students answer questions as often as other students in this school

83. The number of low-income children retained in grade is proportionately equivalent to other children in grade

85. Teachers believe that a student's home background is not the primary factor that determines individual student achievement

100. Students not mastering basic/essential skills are frequently retained in grade

Note. Items 3 and 27 were recoded to match the polarity of other items in this scale.

Table C-5

Opportunity to Learn and Student Time On Task (TOT) Items.

14. Pull out programs (e.g. Chapter I, Special Ed., Gifted, etc.) often disrupt and interfere with basic/essential skills instruction

28. When students are assigned seat work, teachers monitor it closely

30. There are few student interruptions during class time

33. Outside interruptions do not often interfere with basic/essential skill instruction in this school

52. Two hours or more are allocated for reading/language arts each day throughout this school

55. Individual teachers determine allocated time for basic/essential skill instruction without guidelines or discussion from the administration

62. Special instructional programs for individual students are integrated with classroom instruction and the curriculum

65. Typical daily lessons in this school follow this sequence: teacher presentation, student practice, specific feedback, evaluation of student performance

72. Fifty minutes or more is allocated to mathematics instruction each day in this school

76. Teachers in this school expect and plan assignments so that students will be highly successful during practice work following direct instruction

89. Class atmosphere in this school is generally very conducive to learning for all students

98. During basic/essential skills instruction, students are working independently on seat work for the majority of allocated time

---

Note. Items 14, 55, and 98 were recoded to compensate for negative wording.

Table C-6

Frequent Monitoring of Student Progress (FM) Items.

20. Criterion-referenced tests are not used to assess basic skills throughout the school

22. In basic/essential skills instruction in this school, reteaching and specific skills remediation are important parts of the teaching process

37. The standardized testing program is an accurate and valid measure of the basic skills curriculum in this school

40. Teachers and the principal thoroughly review and analyze test results to plan instructional program modifications

42. Standardized test results are not used to evaluate program objectives

43. Multiple assessment methods are used to assess student progress in basic/essential skills (e.g., criterion-referenced tests, work samples, mastery check lists, etc.)

53. Specific feedback on daily assignments is given regularly and followed by the teacher

58. There is no systematic, regular assessment of student's basic/essential skills in most classrooms

61. Student assessment information (such as criterion referenced tests, skills checklists, etc.) is regularly used to give specific feedback and plan appropriate instruction

73. Student assignments in basic/essential skill areas are corrected daily

92. In this school there is annual standardized testing at each grade level

94. Criterion-referenced tests are used to give specific student feedback in basic/essential skills areas throughout the school

---

Note. Items 20, 42, and 58 were coded in reverse order before scales were created.

Table C-7

Positive Home-School Relations (HSR) Items.

9. Most parents would rate this school as superior

31. Homework is monitored at home and in school with follow-up

34. During parent-teacher conferences, there is a focus on factors directly related to student achievement and basic/essential skills mastery

35. Almost all students complete assigned homework before coming to school

36. Ninety to one hundred percent of your students parents attend scheduled parent-teacher conferences

47. There is little cooperation in regard to homework monitoring between parents and teachers in this school

49. Parent-teacher conferences seldom result in specific plans for home-school cooperation aimed at improving student classroom achievement

64. Very few parents of students in your class visit the school to observe the instructional program

67. Most parents understand and promote the school's instructional program

68. There is active parent organization in this school that involves many parent

69. Many parents are involved in an overall home/school support network

71. Teachers and parents are aware of the homework policy in this school

79. Many parents initiate many contacts with this school

84. Phone calls, newsletters, regular notes, and parent-teacher conferences are ways that most teachers communicate with parents in this school

96. Beyond parent conferences and report cards, teachers in this school have several other ways for communicating student progress to parents

---

Note. Responses to items 47, 49, and 64 were recoded to match the polarity of other items in this scales.

APPENDIX D

SCALES RESULTING FROM PRINCIPAL COMPONENTS ANALYSIS

(SEVEN COMPONENT SOLUTION)

Table D-1

Factor Loadings for a Seven Component Solution from South Carolina Data.

#	C-1	#	C-2	#	C-3	#	C-4	#	C-5	#	C-6	#	C-7
				69	.717								
				79	.674								
17	.698			67	.643								
82	.696			68	.638								
48	.662			18	.598	39	.677						
7	.656	53	.594	64	-.566	26	.569						
66	.655	65	.582	36	.536	59	.568						
46	.642	76	.569	5	.513	29	.549			56	.589		
60	.638	73	.553	35	.490	41	.548	49	.522	1	.565		
19	.607	72	.552	47	-.468	54	-.474	42	.515	8	-.520	87	.544
75	.600	61	.528	9	.457	50	.458	58	.506	33	.491	81	.498
78	.583	52	.515	31	.455	11	.409	55	.456	25	.471	95	.423
6	.574	43	.512	74	.387	38	.382	20	.444	24	.441	44	.404
40	.570	99	.503	37	.333	2	.367	15	.413	30	.415	89	.379
21	.566	94	.485	85	<.30	100	.354	32	.398	4	.301	3	<.30
90	.540	34	.481	83	<.30			98	.363	14	<.30		
16	.535	91	.466	80	<.30			27	.350				
10	.526	92	.450					63	.346				
88	-.521	22	.439					12	<.30				
45	.512	84	.438										
57	.504	70	.420										
97	-.478	28	.413										
23	.470	62	.409										
86	.467	96	.408										
93	-.447	77	.384										
13	-.395	71	.325										
		51	< .30										

Note. The column of numbers below each pound sign (#) represents the item number from the Connecticut School Effectiveness Questionnaire (CSEQ). The analysis which provided these factor loadings is based on returns from 103 elementary schools in South Carolina.

Table D-2

<u>Component 1 Items</u>

6. The principal makes several formal classroom observations each year

7. Discussions with the principal often result in some aspect of improved instructional practice

10. The principal reviews and interprets test results with and for the faculty

13. Instructional issues are seldom the focus of faculty meetings

16. The principal uses test results to recommend modifications or changes in the instructional program

17. There is clear, strong, centralized instructional leadership from the principal in this school

19. The principal regularly gives feedback to teachers concerning lesson plans

21. At the principal's initiatives, teachers work together to effectively coordinate the instructional program within and between grades

23. A positive feeling permeates the school

40. Teachers and the principal thoroughly review and analyze test results to plan instructional program modifications

45. The principal requires and regularly reviews lesson plans

46. The principal is very active in securing resources, arranging opportunities and promoting staff development activities for the faculty

48. The principal leads frequent formal discussions concerning instruction and student achievement

57. Formal observations by the principal are regularly followed by a post observation conference

60. The principal frequently communicates to individual teachers their responsibility in relation to student achievement

66. The principal regularly brings instructional issues to the faculty for discussion

Table D-2 (cont'd)

75. The principal is highly visible throughout the school

78. The principal is accessible to discuss matters dealing with instruction

82. The principal is an important instructional resource person in this school

86. During follow-up to formal classroom observations, a plan for improvement frequently results

88. Teachers in this school do not turn to the principal with instructional concerns or problems

90. During follow-up to formal observations, the principal's main emphasis is on instructional improvement

93. The principal rarely makes informal contacts with students and teachers around the school

97. Individual teachers and the principal do not meet regularly to discuss what the principal will observe during a classroom observation

Table D-3

Component 2 Items

22. In basic/essential skills instruction in this school, reteaching and specific skills remediation are important parts of the teaching process

28. When students are assigned seat work, teachers monitor it closely

34. During parent-teacher conferences, there is a focus on factors directly related to student achievement and basic/essential skills mastery

43. Multiple assessment methods are used to assess student progress in basic/essential skills (e.g., criterion-referenced tests, work samples, mastery check lists, etc.)

51. In mathematics, instruction is often presented to homogeneous ability groups

52. Two hours or more are allocated for reading/language arts each day throughout this school

53. Specific feedback on daily assignments is given regularly and followed by the teacher

61. Student assessment information (such as criterion referenced tests, skills checklists, etc.) is regularly used to give specific feedback and plan appropriate instruction

62. Special instructional programs for individual students are integrated with classroom instruction and the curriculum

65. Typical daily lessons in this school follow this sequence: teacher presentation, student practice, specific feedback, evaluation of student performance

70. Supervision is directed at instruction

71. Teachers and parents are aware of the homework policy in this school

72. Fifty minutes or more is allocated to mathematics instruction each day in this school

73. Student assignments in basic/essential skill areas are corrected daily

76. Teachers in this school expect and plan assignments so that students will be highly successful during practice work following direct instruction

Table D-3 (cont'd)

77. Teachers in this school believe they are responsible for all
    students mastering all basic/essential skills at each grade level

84. Phone calls, newsletters, regular notes, and parent-teacher
    conferences are ways that most teachers communicate with parents
    in this school

91. In mathematics, school-wide objectives are the focal point of
    instruction

92. In this school there is annual standardized testing at each grade
    level

94. Criterion-referenced tests are used to give specific student
    feedback in basic/essential skills areas throughout the school

96. Beyond parent conferences and report cards, teachers in this
    school have several other ways for communicating student progress
    to parents

99. In mathematics, written sequential objectives exist in all grades

Table D-4

Component 3 Items

5. Most students in this school are eager and enthusiastic about learning

9. Most parents would rate this school as superior

18. Ninety-five to one hundred percent of the students in this school can be expected to complete high school

31. Homework is monitored at home and in school with follow-up

35. Almost all students complete assigned homework before coming to school

36. Ninety to one hundred percent of your students parents attend scheduled parent-teacher conferences

37. The standardized testing program is an accurate and valid measure of the basic skills curriculum in this school

47. There is little cooperation in regard to homework monitoring between parents and teachers in this school

64. Very few parents of students in your class visit the school to observe the instructional program

67. Most parents understand and promote the school's instructional program

68. There is active parent organization in this school that involves many parent

69. Many parents are involved in an overall home/school support network

74. Teachers believe that all students in this school can master basic/essential skills as a direct result of the instructional program

79. Many parents initiate many contacts with this school

80. Low-achieveing students answer questions as often as other students in this school

83. Th* number of low-income children retained in grade is proportionately equivalent to other children in grade

85. Teachers believe that a student's home background is not the primary factor that determines individual student achievement

Table D-5

Component 4 Items

2. In reading, written sequential objectives exist up through all grades

11. School-wide objectives are the focal point of reading instruction in this school

26. In reading, an identified set of objectives or skills that all students must master exists at each grade level

29. In mathematics, there is an identified set of objectives or skills that all students must master at each grade level

38. Almost all students are expected to master basic/essential skills at each grade level

39. In Language Arts, an identified set of objectives or skills that all students must master exist at each grade level

41. In Language Arts, school-wide objectives are the focus of instruction in this school

50. Reading objectives are coordinated and monitored through all grades

54. In Language Arts, written, sequential objectives do exist at each grade level

59. Language Arts objectives are coordinated and monitored through all grades

100. Students not mastering basic/essential skills are frequently retained in grade

Table D-6

Component 5 Items

12. In reading, initial skill instruction is often presented to a heterogeneous group of students

15. Mathematics objectives are not coordinated and monitored up through all grades in this school

20. Criterion-referenced tests are not used to assess basic skills throughout the school

27. Teachers in this school do not hold consistently high expectations of all students

32. A written statement of purpose that is the driving force behind most important decisions does not exist in this school

42. Standardized test results are not used to evaluate program objectives

49. Parent-teacher conferences seldom result in specific plans for home-school cooperation aimed at improving student classroom achievement

55. Individual teachers determine allocated time for basic/essential skill instruction without guidelines or discussion from the administration

58. There is no systematic, regular assessment of student's basic/essential skills in most classrooms

63. The principal does not put much emphasis on the meaning and use of standardized test results

98. During basic/essential skills instruction, students are working independently on seat work for the majority of allocated time

Table D-7

Component 6 Items

1. This school is a safe and secure place to work

4. Most problems facing this school can be solved by the principal and faculty without a great deal of outside help

8. The physical condition of this school is generally unpleasant and unkempt

14. Pull out programs (e.g. Chapter I, Special Ed., Gifted, etc.) often disrupt and interfere with basic/essential skills instruction

24. All materials and supplies necessary for instruction in basic/essential skills are available

25. Staff and students do not view security as a problem in this school

30. There are few student interruptions during class time

33. Outside interruptions do not often interfere with basic/essential skill instruction in this school

56. The school building in neat, bright, clean and comfortable

Table D-8

<u>Component 7 Items</u>

  3. In this school, low-achieving students present more discipline problems than other students

44. Teachers, administrators and parents assume responsibility for discipline in this school

81. Student behavior is generally positive in this school

87. Students in this school abide by school rules

89. Class atmosphere in this school is generally very conducive to learning for all students

95. Generally, discipline is not a problem in this school

APPENDIX E

SCALES DERIVED FROM PRINCIPAL COMPONENTS ANALYSIS

(SIX COMPONENT SOLUTION)

Table E-1

Factor Loadings from Six Component Solution for CSEQ Data.

#	C-3	#	C-5	#	C-4	#	C-2	#	C-1
82	.705								
17	.699								
7	.658			69	.716	39	.676		
48	.655	53	.591	79	.670	26	.579		
66	.649	65	.574	67	.641	29	.565		
46	.640	76	.557	68	.640	59	.554		
60	.632	72	.547	18	.592	41	.539	1	.566
75	.612	73	.539	64	-.570	54	-.457	56	.545
19	.601	61	.523	36	.522	50	.454	87	.530
78	.589	52	.512	5	.512	38	.406	81	.517
6	.569	43	.502	35	.492	11	.405	8	-.514
21	.567	99	.491	47	-.473	100	.392	95	.497
40	.558	94	.484	31	.449	2	.355	25	.433
88	-.540	34	.478	9	.447			24	.429
90	.539	91	.459	74	.394			33	.417
16	.521	92	.454	71	.329			41	.412
10	.515	84	.432	37	.321			4	.412
45	.506	22	.428	85	<.30			30	.396
57	.489	70	.414	83	<.30			89	.386
97	-.482	28	.407	80	<.30				
23	.479	96	.405	3	<.30				
93	-.470	62	.400						
86	.466	77	.366						
13	-.393	51	<.30						
		14	<.30						

Note. The column of numbers below each pound sign (#) represents the item number from the Connecticut School Effectiveness Questionnaire (CSEQ). Underlined items were deleted from these factors to arrive at the final set of scales used in subsequent analyses. The analysis which provided these factor loadings is based on returns from 103 elementary schools in South Carolina.

Table E-2

Items Corresponding to Component 3.

6. The principal makes several formal classroom observations each year

7. Discussions with the principal often result in some aspect of improved instructional practice

10. The principal reviews and interprets test results with and for the faculty

16. The principal uses test results to recommend modifications or changes in the instructional program

17. There is clear, strong, centralized instructional leadership from the principal in this school

19. The principal regularly gives feedback to teachers concerning lesson plans

21. At the principal's initiatives, teachers work together to effectively coordinate the instructional program within and between grades

45. The principal requires and regularly reviews lesson plans

46. The principal is very active in securing resources, arranging opportunities and promoting staff development activities for the faculty

48. The principal leads frequent formal discussions concerning instruction and student achievement

57. Formal observations by the principal are regularly followed by a post observation conference

60. The principal frequently communicates to individual teachers their responsibility in relation to student achievement

66. The principal regularly brings instructional issues to the faculty for discussion

75. The principal is highly visible throughout the school

78. The principal is accessible to discuss matters dealing with instruction

82. The principal is an important instructional resource person in this school

Table E-2 (cont'd)

86. During follow-up to formal classroom observations, a plan for improvement frequently results

88. Teachers in this school do not turn to the principal with instructional concerns or problems

90. During follow-up to formal observations, the principal's main emphasis is on instructional improvement

93. The principal rarely makes informal contacts with students and teachers around the school

97. Individual teachers and the principal do not meet regularly to discuss what the principal will observe during a classroom observation

---

Note. This scale parallels the Strong Instructional Leadership (IL) construct.

Table E-3

Items Corresponding to Component 5.

28. When students are assigned seat work, teachers monitor it closely

52. Two hours or more are allocated for reading/language arts each day throughout this school

62. Special instructional programs for individual students are integrated with classroom instruction and the curriculum

65. Typical daily lessons in this school follow this sequence: teacher presentation, student practice, specific feedback, evaluation of student performance

72. Fifty minutes or more is allocated to mathematics instruction each day in this school

76. Teachers in this school expect and plan assignments so that students will be highly successful during practice work following direct instruction

Note. Items in this scale generally cluster under the "Frequent Monitoring of Student Progress (FM)" construct.

Table E-4

Items Clustering under Component 4.

9. Most parents would rate this school as superior

31. Homework is monitored at home and in school with follow-up

35. Almost all students complete assigned homework before coming to school

36. Ninety to one hundred percent of your students parents attend scheduled parent-teacher conferences

47. There is little cooperation in regard to homework monitoring between parents and teachers in this school

64. Very few parents of students in your class visit the school to observe the instructional program

67. Most parents understand and promote the school's instructional program

68. There is active parent organization in this school that involves many parent

69. Many parents are involved in an overall home/school support network

71. Teachers and parents are aware of the homework policy in this school

79. Many parents initiate many contacts with this school

Note. Both the High Expectations and Positive Home-School Relations constructs are represented in this scale.

Table E-5

Items Loading on Component 2.

2. In reading, written sequential objectives exist up through all grades

11. School-wide objectives are the focal point of reading instruction in this school

26. In reading, an identified set of objectives or skills that all students must master exists at each grade level

29. In mathematics, there is an identified set of objectives or skills that all students must master at each grade level

39. In Language Arts, an identified set of objectives or skills that all students must master exist at each grade level

41. In Language Arts, school-wide objectives are the focus of instruction in this school

50. Reading objectives are coordinated and monitored through all grades

54. In Language Arts, written, sequential objectives do exist at each grade level

59. Language Arts objectives are coordinated and monitored through all grades

___

Note. All items comprising this factor are subsumed by the Clear School Mission construct.

Table E-6

<u>Items Associated with Component 1</u>.

1. This school is a safe and secure place to work

8. The physical condition of this school is generally unpleasant and unkempt

44. Teachers, administrators and parents assume responsibility for discipline in this school

56. The school building in neat, bright, clean and comfortable

81. Student behavior is generally positive in this school

87. Students in this school abide by school rules

95. Generally, discipline is not a problem in this school

<u>Note</u>. The SAFE factor items cluster primarily within the Positive School Climate construct.

APPENDIX F

FINAL SCHOOL EFFECTIVENESS SCALES

Table F-1

ORDERLY ENVIRONMENT Items.

1. This school is a safe and secure place to work

8. The physical condition of this school is generally unpleasant and unkempt

56. The school building in neat, bright, clean and comfortable

81. Student behavior is generally positive in this school

87. Students in this school abide by school rules

95. Generally, discipline is not a problem in this school

Note. The coding for item number 8 was reversed to match the response set of other scale items. Mean - 3.82, SD - .47, Range 2.62 to 4.58.

Table F-2

INSTRUCTIONAL OBJECTIVES Items.

2. In reading, written sequential objectives exist up through all grades

11. School-wide objectives are the focal point of reading instruction in this school

26. In reading, an identified set of objectives or skills that all students must master exists at each grade level

29. In mathematics, there is an identified set of objectives or skills that all students must master at each grade level

39. In Language Arts, an identified set of objectives or skills that all students must master exist at each grade level

41. In Language Arts, school-wide objectives are the focus of instruction in this school

50. Reading objectives are coordinated and monitored through all grades

54. In Language Arts, written, sequential objectives do exist at each grade level

59. Language Arts objectives are coordinated and monitored through all grades

---

Note. The coding for item number 54 was reversed to match the rest of the scale. Mean - 4.06, SD - .35, Range 1.99 to 4.70.

Table F-3

PRINCIPAL LEADERSHIP Items.

6. The principal makes several formal classroom observations each year

7. Discussions with the principal often result in some aspect of improved instructional practice

10. The principal reviews and interprets test results with and for the faculty

16. The principal uses test results to recommend modifications or changes in the instructional program

17. There is clear, strong, centralized instructional leadership from the principal in this school

19. The principal regularly gives feedback to teachers concerning lesson plans

21. At the principal's initiative, teachers work together to effectively coordinate the instructional program within and between grades.

45. The principal requires and regularly reviews lesson plans

46. The principal is very active in securing resources, arranging opportunities and promoting staff development activities for the faculty

48. The principal leads frequent formal discussions concerning instruction and student achievement

57. Formal observations by the principal are regularly followed by a post observation conference

60. The principal frequently communicates to individual teachers their responsibility in relation to student achievement

66. The principal regularly brings instructional issues to the faculty for discussion

75. The principal is highly visible throughout the school

78. The principal is accessible to discuss matters dealing with instruction

82. The principal is an important instructional resource person in this school

Table F-3 (cont'd)

86. During follow-up to formal classroom observations, a plan for improvement frequently results

88. Teachers in this school do not turn to the principal with instructional concerns or problems

90. During follow-up to formal observations, the principal's main emphasis is on instructional improvement

93. The principal rarely makes informal contacts with students and teachers around the school

97. Individual teachers and the principal do not meet regularly to discuss what the principal will observe during a classroom observation

---

Note.   Responses to items 88, 93, and 97 were recoded to match the polarity of other items in the PRN scale. The scale was then formed by adding item scores together (with unit weighting) and dividing by 21 (i.e., the number of items in the scale). Mean - 3.75, SD - .43, Range 2.03 to 4.60

Table F-4

MONITORING & FEEDBACK Items.

22. In basic/essential skills instruction in this school, reteaching and specific skills remediation are important parts of the teaching process

28. When students are assigned seat work, teachers monitor it closely

43. Multiple assessment methods are used to assess student progress in basic/essential skills (e.g., criterion-referenced tests, work samples, mastery check lists, etc.)

52. Two hours or more are allocated for reading/language arts each day throughout this school

53. Specific feedback on daily assignments is given regularly and followed by the teacher

62. Special instructional programs for individual students are integrated with classroom instruction and the curriculum

65. Typical daily lessons in this school follow this sequence: teacher presentation, student practice, specific feedback, evaluation of student performance

72. Fifty minutes or more is allocated to mathematics instruction each day in this school

73. Student assignments in basic/essential skill areas are corrected daily

76. Teachers in this school expect and plan assignments so that students will be highly successful during practice work following direct instruction

92. In this school there is annual standardized testing at each grade level

94. Criterion-referenced tests are used to give specific student feedback in basic/essential skills areas throughout the school

---

Note. Mean - 4.11 SD - .26, Range 2.77 to 4.64.

Table F-5

PARENT INVOLVEMENT Items.

9. Most parents would rate this school as superior

31. Homework is monitored at home and in school with follow-up

35. Almost all students complete assigned homework before coming to school

36. Ninety to one hundred percent of your students parents attend scheduled parent-teacher conferences

47. There is little cooperation in regard to homework monitoring between parents and teachers in this school

64. Very few parents of students in your class visit the school to observe the instructional program

67. Most parents understand and promote the school's instructional program

68. There is active parent organization in this school that involves many parent

69. Many parents are involved in an overall home/school support network

71. Teachers and parents are aware of the homework policy in this school

79. Many parents initiate many contacts with this school

Note. Responses to items 47 and 64 were recoded to match the polarity of other items in this scale. The scale was then formed by adding item scores together (with unit weighting) and dividing by the number of items in the HSR scale (i.e., 11).
Mean = 3.75, SD = .40, Range 2.32 to 4.17.