A New Era of School Reform: Going Where the Research Takes Us

REL Contract #RJ96006101
Deliverable 2000-05

prepared by

Robert J. Marzano

Mid-continent Research for Education and Learning
2550 S. Parker Road, Suite 500
Aurora, CO 80014
303-337-0990 (phone)
303-337-3005 (fax)

To order copies of *A New Era of School Reform: Going Where the Research Takes Us*, contact McREL:

Mid-continent Research for Education and Learning
2550 S. Parker Road, Suite 500
Aurora, Colorado 80014
tel: 303-337-0990
fax: 303-337-3005
web site: mcrel.org
e-mail: info@mcrel.org

This work was produced in whole or in part with funds from the Office of Educational Research and Improvement (OERI), U.S. Department of Education, under Contract Number RJ96006101. The content does not necessarily reflect the views of OERI or the Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the federal government.

TABLE OF CONTENTS

Chapter 1
A QUESTION OF SCHOOLING

As the title indicates, the central thesis of this monograph is that educators stand at the dawn of a new era of school reform. This is not because a new decade, century, and millennium are beginning, although these certainly are noteworthy events. Rather, it is because the cumulative research of the last 40 years provides some clear guidance about the characteristics of effective schools and effective teaching. Knowledge of these characteristics provides educators with possibilities for reform unlike those available at any other time in history. In fact, one of the primary goals of this monograph is to synthesize that research and translate it into principles and generalizations educators can use to effect substantive school reform.

The chapters that follow attempt to synthesize and interpret the extant research on the impact of schooling on students' academic achievement. The interval of four decades has been selected because this is the period during which the effects of schooling have been systematically studied. According to Madaus, Airasian, and Kellaghan (1980):

> In the 1950s and early 1960s, the struggle against poverty, racial and unequal educational opportunity became more intense. Starting just after 1960, the effort to deal with these problems dominated domestic legislative action. . . . Attempts to document and remedy the problems of unequal educational opportunity, particularly as they related to minority-group children, provided the major impetus for school-effectiveness studies. In fact, major societal efforts to address the problems of inequality were centered on the educational sphere. (p. 11)

It was in this context that the Civil Rights Act of 1964, a cornerstone of President Johnson's "war on poverty," specified that the Commissioner of Education should conduct a nationwide survey of the availability of educational opportunity. The wording of the mandate revealed an assumption on the part of the Act's authors that educational opportunity was not equal for all members of American society:

> The Commissioner shall conduct a survey and make a report to the President and Congress. . .concerning the *lack of availability of equal educational opportunities* [emphasis added] for individuals by reason of race, color, religion, or national origin in public institutions. (In Madaus, Airasian, & Kellaghan, 1980, p. 12)

Madaus, Airasian, and Kellaghan explain: "It is not clear why Congress ordered the commissioner to conduct the survey, although the phrase 'concerning the lack of availability of educational opportunities' implies that Congress believed that inequalities in opportunities did exist, and that documenting these differences could provide a useful legal and political tool to overcome future oppositions to school reform" (p. 12). According to Mosteller and Moynihan (1972), James Coleman, who was selected to head the team of researchers conducting the survey, indicated in an interview that he believed the study would disclose a great disparity in the quality of education afforded black versus white students — a fact interpreted by Mosteller and Moynihan as evidence that Coleman began the study with a conclusion already in mind.

Whether the project was undertaken with a bias has always been and will continue to be a matter of speculation only. However, it is not a matter of speculation that the study was the largest survey of public education ever undertaken. Over 640,000 students in grades 1, 3, 6, 9, and 12 categorized into six ethnic and cultural groups took achievement tests and aptitude tests. About 60,000 teachers in over 4,000 schools completed questionnaires about their background and training.

The report, published in July 1966, is entitled *Equality of Educational Opportunity* but commonly is referred to as the "Coleman Report" in deference to its senior author. The findings were not favorable regarding the impact of schooling:

> Taking all of these results together, one implication stands above all: that schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school. (p. 325)

Madaus et al. (1980) explain that the report had two primary effects on perceptions about schooling in America. First, it dealt a blow to the perception that schools could be a viable agent in equalizing the disparity in students' academic achievement due to environmental factors. Second, it spawned the perception that differences in schools have little, if any, relationship with student achievement. One of the most well-publicized findings from the report was that schools account for only about 10 percent of the variances in student achievement — the other 90 percent was accounted for by student background characteristics.

Coleman et al.'s findings were corroborated in 1972 when Jencks and his colleagues (1972) published *Inequality: A Reassessment of the Effects of Family and Schooling in America*, which was based on a re-analysis of data from the Coleman report. Among the findings articulated in the Jencks study were the following:

- Schools do little to lessen the gap between rich and poor students.
- Schools do little to lessen the gap between more and less abled students.
- Student achievement is primarily a function of one factor — the background of the student.
- There is little evidence that education reform can improve the influence a school has on student achievement.

Taken at face value, the conclusions articulated and implied in the Coleman and Jencks reports paint a somber picture for education reform. If schools have little chance of overcoming the influence of students' background characteristics, why put any energy into school reform?

More than three decades have passed since the commissioned survey was undertaken. What have we learned since then? Is the picture of schooling more positive now? This monograph attempts to answer these questions. As the following chapter will illustrate, when the research undertaken during the last four decades is considered as a set, there is ample evidence that schools can and do make a powerful difference in the academic achievement of students.

A NECESSARILY TECHNICAL LOOK

The discussion in this monograph is somewhat technical in nature. This is necessarily the case because the research on school effectiveness has become quite sophisticated, both in terms of methodology and statistics, particularly over the last two decades. (For a discussion of these changes, see Willms, 1992; Byrk & Raudenbush, 1992.) However, an attempt has been made to include discussions of formulae and the rationale for specific data analysis and estimation techniques used in this monograph. These explanations can be found in footnotes and, where appropriate, in endnotes after each chapter.

Throughout this monograph, five indices are used to describe the relationship between student achievement and various school-, teacher-, and student-level factors.

Percent of Variance Explained: *PV*

One of the most common indices found in the research on the effects of schooling is the percent of variance explained, or *PV* as referred to in this monograph. As mentioned previously, this was the index used by Coleman for interpreting the survey data. A basic assumption underlying the use of this index is that the percent of variance explained by a predictor or independent variable (e.g., schooling) relative to a predicted or dependent variable (e.g., student achievement) is a good indication of the strength of relation between the two. Most commonly, a "set" of predictor variables is used. For example, a given study might attempt to predict student achievement using (1) per-pupil expenditures, (2) proportion of academic classes, and (3) average years of experience per teacher. The predictor variables considered as a set would account for a proportion of total variance in the predicted variable[1]. The index used to judge the influence of predictor variables is the ratio of variance accounted for by the predictor variables over the total variance of the predicted variable multiplied by 100. As mentioned previously, this index is referred to in this monograph as *PV*:

$$PV = \frac{\text{percent of variance explained by predictor or independent variables}}{\text{percent of total variance in the predicted or dependent variable}} \times 100$$

The Correlation Coefficient: *r* and *R*

An index closely related to *PV* is the correlation coefficient. When a single predictor or independent variable (e.g., socioeconomic status) is used with a predicted or dependent variable (e.g., students' academic achievement), the relationship between the two is expressed as r — the Pearson product-moment correlation. When multiple predictors (e.g., prior knowledge, quality of the school,

[1]The process of determining the relationship between a predicted or dependent variable and predictor or independent variables is commonly referred to as "regression analysis." The predictor variable is "regressed onto" the predictor variable. The reader will note that this phrase is used frequently throughout the monograph.

socioeconomic status) are used with a predicted variable, the relationship between the predictor variables considered as a set and the predicted variable is expressed as R — the multiple correlation coefficient. In both cases, the percent of variance accounted for (*PV*) in the predicted or dependent variable by the predictor or independent variables is computed by squaring the correlation coefficient (i.e., r^2 or R^2) and multiplying by 100. In short, there is a strong conceptual and mathematical relationship between *PV* and the univariate and multi-variate correlation coefficients. Commonly, when school effects are expressed in one metric, they are also expressed in the other.

As common as is the use of these metrics, they have been criticized as indicators of the relationship between predictor or independent and predicted or dependent variables in the research on school effectiveness. This is especially the case with *PV*, as Hunter and Schmidt (1990) explain:

> The percent of variance accounted for is statistically correct, but substantively erroneous. It leads to severe underestimates of the practical and theoretical significance of relationships between variables. . . .The problem with all percent variance accounted for indices of effect size is that variables that account for small percentages of the variance often have very important effects on the dependent variable. (pp. 199–200)

To illustrate this circumstance, Hunter and Schmidt use the correlation between aptitude and heredity reported by Jensen (1980). This correlation is about .895, which implies that about 80 percent ($.895^2$) of the (true) variance in aptitude is a function of heredity, leaving only 20 percent of the variance due to environment (r = .447). The relative influence of heredity on aptitude, and environment on aptitude, then, is about 4 to 1 from the percent of variance perspective. However, regression theory (see Cohen & Cohen, 1975) tells us that the correlations between heredity and aptitude (H) and between environment and aptitude (E) (after the influence of heredity has been partialed out) are analogous to the regression weights in a linear equation predicting aptitude from heredity and environment when dependent and independent variables are expressed in standard score form. (For this illustration, we will assume that heredity and environment are independent.) Using the quantities above, this equation would be as follows:

$$\text{Predicted Aptitude} = .895(H) + .447(E)$$

This equation states that an increase of one standard deviation in heredity will be accompanied by an increase of .895 standard deviations in aptitude. Similarly, an increase of one standard deviation in environment will be accompanied by an increase of .447 standard deviations in aptitude. This paints a very different picture of the relative influences of heredity and environment on aptitude. Here the ratio is 2 to 1 as opposed to 4 to 1 from the percent of variance perspective.

The Binomial Effect Size Display: *BESD*

The potentially misleading impressions given by the correlation coefficient and the percent of variance explained has stimulated the use of a third metric — the binomial effect size display (*BESD*). Rosenthal and Rubin (1982) explain that the percent of variance accounted for index invites misleading interpretations of the relative influence of predictor variables on predicted variables. Whereas r or R can be interpreted with distortion (as evidenced above), the *BESD* provides for the most useful interpretation. The *BESD* is similar to the interpretation one would use with a fourfold (tetrachoric or phi) correlation coefficient[2]. Rosenthal and Rubin explain that most education studies can be conceptualized this way by dichotomizing the predictor or independent variable (membership in either the experimental or control group) and the predicted or dependent variable (success or failure on the criterion measure). Using these dichotomies, the *BESD* allows for interpretation of comparative success or failure on the criterion as a function of membership in an experimental or control group. Cohen (1988) dramatically illustrates the utility of the *BESD* using an example from medicine. (See Table 1.1.)

Table 1.1
Binomial Effect Size Display With 1% of Variance (r = .10) Accounted For
Effects of Hypothetical Medical Treatment

Group	Outcome%		
	%Alive	%Dead	Total
Treatment	55%	45%	100%
Control	45%	55%	100%

Note: Constructed from data in *Statistical Power for the Behavioral Sciences*, p. 534, by J. Cohen, 1988, Hillsdale, NJ: Erlbaum. r stands for the Pearson product-moment correlation coefficient. See note at the end of Chapter 3 for more information about this quantity.

Table 1.1 exemplifies a situation in which the independent variable (i.e., membership in the experimental or control group) accounts for only one percent of the variance in the dependent variable (i.e., r = .10). The assumption here is that the independent variable is some sort of medical treatment that accounts for one percent of the variance in the outcome measure, which is being alive or dead. Yet, this one percent of explained variance translates into a 10 percentage-point difference in terms of patients who are alive (or dead) based on group membership. As Cohen (1988) notes:

[2]A fourfold or tetrachoric correlation is basically equivalent to a Pearson product-moment correlation (r) when both the predictor variable and the predicted variable are dichotomized. Relative to the *BESD*, the predictor variable is thought of as being dichotomized into two distinct groups. In most of the *BESD* illustrations used in this monograph, the dichotomized independent variable will be thought of as effective schools versus ineffective schools. Similarly, relative to the *BESD*, the predicted variable is dichotomized into success or failure on some criterion measure. In this monograph, the predicted variable will generally be thought of as success or failure on some form of achievement test.

A common convention when using the *BESD* is to assume that the expectation for the predicted variable is a success rate of .50. To compute the *BESD*, the correlation coefficient is divided by 2 and then added to and subtracted from .50. For example, if the r between predictor and predicted is .50, then .50 ÷ 2 = .25. The percentage of subjects in the experimental group that would be expected to “succeed” on the predicted variable is computed as .50 + .25 = .75. The percentage of subjects in the experimental group that would be expected to “fail” on the criterion measure is .50 − .25 = .25. The converse of these computations is used for the control group. Rosenthal and Rubin (1982) make the case for the use of *BESD* as a realistic representation of the size of the treatment effect when the outcome variable is continuous, provided that the groups are of equal size and variance.

> This means, for example, that a difference in percent alive between .45 and .55, which most people would consider important (*alive*, mind you!) yields $r = .10$, and "only 1% of the variance accounted for," an amount that operationally defines a "small" effect in my scheme. . . .
>
> "Death" tends to concentrate the mind. But this in turn reinforces the principle that the size of an effect can only be appraised in the context of the substantive issues involved. An r^2 of .01 is indeed small in absolute terms, but when it represents a ten percentage point increase in survival, it may well be considered large. (p. 534)

This same point is further dramatized by Abelson (1985). After analyzing the effect of various physical skills on the batting averages of professional baseball players, he found that the percent of variance accounted for by these skills was a minuscule .00317 — not quite one-third of one percent ($r = .056$). Commenting on the implications for interpreting education research, Abelson notes:

> One should not necessarily be scornful of minuscule values for percentage of variance explained, provided there is statistical assurance that these values are significantly above zero, and that the degree of potential cumulation is substantial. (p. 133)

Finally, Cohen exhorts: "The next time you read 'only X% of the variance is accounted for,' remember Abelson's paradox" (p. 535).

The *BESD* provides for an interesting perspective on the findings from the Coleman report — namely, that schooling accounts for only about 10 percent of the variance in student achievement. When the associated r of .316 is displayed in terms of the *BESD*, the results lead to a different interpretation than that promoted by Coleman. This is shown in Table 1.2. To interpret Table 1.2, assume that the criterion measure is a state test that 50 percent of students are expected to pass.

As illustrated in Table 1.2, when the 10 percent of the variance in student achievement accounted for by schooling is thought of in terms of success or failure on some measure (e.g., a state test on standards), the difference between "effective" and "ineffective" schools is dramatic. Specifically, 31.6 percent more students would pass the test in effective schools than in ineffective schools.

Table 1.2
Binomial Effect Size Display with 10% of Variance ($r = .316$) Accounted For

Group	**Outcome%**		
	% Success	% Failure	Total
Effective Schools	65.8%	34.2%	100%
Ineffective Schools	34.2%	65.8%	100%

The Standardized Mean Difference Effect Size: *ESd*

Another index commonly used in discussions of the effects of schooling is the standardized mean difference. Glass (1976) first popularized this index now commonly used in research on school effects. Commonly referred to as an effect size[3], the index is the difference between experimental and control means divided by an estimate of the population standard deviation — hence, the name, standardized mean difference.

$$\text{standardized mean difference effect size} = \frac{\bar{x}\ \text{experimental group} - \bar{x}\ \text{control group}}{\text{estimate of population standard deviation}}$$

Theorists have suggested a variety of ways to estimate the population standard deviation along with techniques for computing the effect size index under different assumptions (see Cohen, 1988; Glass, 1976; Hedges and Olkin, 1985). The effect size index used throughout this monograph uses the pooled standard deviation from experimental and control groups as the population estimate. It is frequently referred to as Cohen's *d*. It will be referred to as *ESd* throughout the remainder of this monograph.

To illustrate the use of *ESd*, assume that the achievement mean of a school with a given characteristic is 90 on a standardized test and that the mean of a school that does not possess this characteristic is 80. Also assume that the population standard deviation is 10. The effect size would be

$$ESd = \frac{90 - 80}{10} = 1.0$$

This effect size can be interpreted in the following way: the mean of the experimental group is 1.0 standard deviations larger than the mean of the control group. One might infer, then, that the characteristic possessed by the experimental school raises achievement test scores by one standard deviation. Thus, the effect size (*ESd*) expresses the differences between means in standardized or Z score form[4]. It is this characteristic that gives rise to the fifth index commonly used in the research on school effects — percentile gain.

Percentile Gain: *P gain*

Percentile gain (*P gain*) is the expected gain (or loss) in percentile points of the average student in the experimental group compared to the average student in the control group. To illustrate, consider the example above. Given an effect size, *ESd*, of 1.0, one can conclude that the average score in the

---

[3]In this monograph, the term "effect size" and its related symbol *ESd* are reserved for the standardized mean difference. However, it is important to note that *r*, *R*, and *PV* are also referred to as effect sizes in the literature.

[4]Z scores are standardized scores with a mean of 0 and a standard deviation of one.

experimental group is 34.134 percentile points higher than the average score in the control group. This is necessarily so since the *ESd* translates the difference between experimental and control group means into Z score form. Distribution theory tells us that a Z score of 1.0 is at the 84.134 percentile point of the standard normal distribution. To compute the *P gain*, then, *ESd* is transformed into percentile points above or below the 50^{th} percentile point on the standard normal distribution.

The Five Indices

In summary, five indices are commonly used in the research on school effects and form the basis for the discussion to follow. As used in this monograph, those indices are *PV*, *r* or *R*, *BESD*, *ESd*, and *P gain*. Table 1.3 provides the explanations for these indices and their relationships.

These indices are used somewhat interchangeably throughout this monograph. The reader is cautioned to keep in mind the preceding discussion about the characteristics of each index and their interpretations and possible misinterpretations. The selection of the most appropriate indices to use in the following discussion was based on the indices used in the original research and the appropriateness of the indices to the overall point of the discussion.

PURPOSE AND DIRECTION OF THIS MONOGRAPH

As the previous discussion indicates, there are many ways to analyze and interpret the research on school effects. One basic question addressed in this report is whether the 30-plus years of research since the Coleman report still supports the finding that schooling accounts for only 10 percent of variance in student achievement. A second basic question addressed is, What are the school-, classroom-, and student-level factors that influence student achievement?

Limitations

It should be noted at the outset that this monograph focuses only on those school- and teacher-level characteristics that can be implemented without drastic changes in resources or personnel. By definition, then, interventions that would require exceptional resources (e.g., year-round school, computers for every student, after-school programs) or additional personnel (e.g., lower teacher/student ratios, tutoring for students) are not addressed in this report. This is not to say that these are not viable reform efforts. Indeed, structural changes such as these might hold the ultimate solution to school reform. However, this report focuses on changes that can be implemented given the current structure and resources available to schools.

Outline

The remaining chapters in this monograph are organized in the following manner. The first section, "Part I: General Literature Review," includes Chapters 2 and 3, which review the literature on previous attempts to identify those variables impacting student achievement. Chapter 2 focuses on studies that were part of the "school effectiveness movement"; Chapter 3 focuses on studies that were not part of this movement and that were more synthetic in nature. The studies in Chapter 3 might be considered "classic" studies of the effects of schooling. The second section, "Part II: Research on School, Teacher, and Student Effects," includes Chapters 4, 5, and 6. Chapter 4 presents a discussion of the research on school-level variables. Chapters 5 and 6, respectively, review the research on teacher-level variables and student-level variables. The final section, "Part III: Applications," includes Chapter 7, which considers the implications of the findings from Chapters 4, 5, and 6 for school reform.

Table 1.3
Indices Used in This Monograph

Symbol	Name	Explanation and Relationship to Other Indices
PV	percent of variance explained	Percentage of variance in the predicted or dependent variable accounted for or explained by the predictor or independent variables. *PV* is commonly computed by squaring r (when one predictor or independent variable is involved) or squaring R (when multiple predictors or independent variables are involved).
r or R	bivariate correlation coefficient and multiple correlation coefficient	Relationship between predictor(s) and predicted variable expressed as an index from −1.0 to +1.0 in the case of r, and .00 to +1.00 in the case of R. r^2 and R^2 are equivalent to *PV*. When one independent or predictor variable is involved, *ESd* is equal to $2r/\sqrt{1-r^2}$.
BESD	binomial effect size display	The expected difference between experimental and control groups relative to the percentage of students who would pass a test on which the normal passing rate is 50%. *BESD* is usually computed using r. Specifically, $r/2$ is added and subtracted from 50%.
ESd	standardized mean difference effect size	The difference between the experimental group mean and the control group mean standardized by an estimate of the population standard deviation. *ESd* can be converted to r via the following formula: $$r = \frac{ESd}{\sqrt{ESd^2 + 4}}$$
P gain	percentile gain	The difference in percentile points between the mean of the experimental group and the mean of the control group. *P gain* is computed by transforming *ESd* to a percentile point in the standard normal distribution and then subtracting 50%.

PART I:
GENERAL LITERATURE REVIEW

Chapter 2
THE SCHOOL EFFECTIVENESS MOVEMENT

There was a rather swift reaction to the works of Coleman and Jencks from the world of education research. A number of efforts were launched to demonstrate the effectiveness of schools and to rather pointedly provide a counter argument to that implicit in the Coleman and Jencks studies. This chapter reviews studies that fall into the category of what might loosely be referred to as the "school effectiveness movement."

Arguably, the school effectiveness movement can be thought of as a set of studies and reform efforts that took place in the 1970s and early 1980s and shared the common purpose of identifying those within-school factors that affect students' academic achievement. The case might also be made that studies in this category were loosely joined by virtue of the people conducting the studies (i.e., a relatively small network of like-minded researchers) and/or by antecedent/consequent relationships between studies (i.e., one study built on the findings from a previous study). (For an extensive review of the school effectiveness research, see Good and Brophy, 1986.)

EDMONDS

It is probably accurate to say that Ron Edmonds is considered the figurehead of the school effectiveness movement. As Good and Brophy (1986) note:

> Until his untimely death in 1983, [Edmonds] had been one of the key figures in the school effectiveness movement. . . . Edmonds, more than anyone, had been responsible for the communication of the belief that *schools* can and do make a difference. (p. 582)

Edmonds' contributions were primarily provocative and conceptual in nature (see Edmonds, 1979a, 1979b, 1979c, 1981a, 1981b; Edmonds & Frederiksen, 1979). First and foremost, Edmonds asserted that schools can and do make a difference in student achievement. In addition, he operationalized the definition of effective schools as those that close the achievement gap between students from low socioeconomic (SES) backgrounds and those from high socioeconomic backgrounds. Perhaps his most salient contribution was the articulation of the five "correlates" — five school-level variables that allegedly are strongly correlated with student achievement:

1. Strong administrative leadership
2. High expectations for student achievement
3. An orderly atmosphere conducive to learning
4. An emphasis on basic skill acquisition
5. Frequent monitoring of student progress

Although other researchers proposed somewhat different lists (see Purkey & Smith, 1982, for a discussion), Edmonds' five correlates of effective schools became immensely popular. As Scheerens and Bosker (1997) explain, these five correlates became the framework for thinking about school effectiveness for at least a decade, although probably longer.

RUTTER

Concomitant with Edmonds' work was Rutter's study of secondary students in London, which culminated in the popular book *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children* (Rutter, Maughan, Mortimer, & Ouston, 1979). Rutter et al. used what might be loosely referred to as a longitudinal design. In a previous study in 1970, all ten-year-olds in one London borough were tested on general aptitude, reading achievement, and behavioral problems. In 1974, Rutter followed up on students in this cohort group who attended 20 nonselective secondary schools. Students were again tested for aptitude, reading achievement, and behavioral problems. Demographic data also were collected on each student relative to home environment, parental education, level of income, and the like. These data were used as baseline "intake" data to control for student differences. In 1976, students were again assessed in four general areas: attendance, behavior, academic achievement, and delinquency. In addition, the schools they attended were studied relative to a number of school-level variables. The 1976 outcome measures for students were then corrected or adjusted using the intake data, and schools were ranked on the various outcome measures. Rank-order correlations were computed between school characteristics and school rank on the various outcome measures. Some of the more salient findings as reported by Rutter et al. are summarized in Table 2.1.

Table 2.1
Findings from the Rutter et al. Study

Schools differed significantly in the behavioral problems even after correcting for the intake behavioral characteristics of their students.

Schools differed in their corrected verbal reasoning.

Schools' physical and material characteristics had little or no relationship with the behavior of students or their academic achievement.

Characteristics that correlated positively with student behavior were

- attention to homework,
- total teaching time per week,
- class lesson preparation,
- positive expectations, and
- positive reward was generally more effective than negative reward.

Process variables that had a significant relationship with student outcome measures were

- academic emphasis,
- teaching behavior,
- use of reward and punishment,
- degree of student responsibility,
- staff stability, and
- staff organization.

Note: See *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*, by M. Rutter, B. Maughan, P. Mortimer, and J. Ouston, 1979, London: Open Books.

One aspect of the Rutter study that complicated the interpretation of its findings was the use of rank-order correlations. This statistic does not allow for a straightforward interpretation of the strength of relationships between student achievement and the various outcome measures, such as *ESd* or *PV*, for at least two reasons. First, the unit of analysis is the school. Consequently, within-school variance due to differences between individual students is not analyzed. Second, the magnitude of differences between schools is lost with rank-order correlations. In fact, when a straightforward, multiple-regression analysis was performed using individual student achievement as the dependent variable, and student aptitude, parental occupation, selected SES factors, and school process as the independent variables, school process variables uniquely accounted for only 1.6 percent of the total variance. In spite of its shortcomings, the publication of *15,000 Hours* had a powerful effect on school reform efforts in Britain and the United States, sparking intense interest in the study of effective schools.

KLITGAARD AND HALL

Klitgaard and Hall's (1974) study was arguably the first, rigorous, large-scale attempt to identify variables associated with effective schools (Good & Brophy, 1986). These researchers analyzed three sets of data: two years' worth of scores from 4th and 7th graders from 90 percent of Michigan schools, achievement scores from grades 2–6 in New York City, and scores from the Project Talent high school study. After analyzing residual scores from the regression of achievement scores on student background variables, they concluded that of the 161 Michigan schools in the study, about nine percent (i.e., 15) increased student achievement by one standard deviation (i.e., had an *ESd* of 1.0) after controlling for background variables. Similarly, of the 627 schools in the New York sample, the residual achievement of 30 schools was one standard deviation above the mean.

Although the Klitgaard and Hall study provided clear evidence that some schools produce relatively large gains in student achievement, these "high-achieving" schools represented a small minority of those in the population. In addition, the Klitgaard and Hall study did not address whether the "highly effective schools" were equally effective for students from all backgrounds.

BROOKOVER ET AL.

The study by Brookover and his colleagues (Brookover et al., 1978; Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979) was one of the most significant school effectiveness studies, not only for its timing (i.e., it was one of the early studies conducted on school-level variables), but also for its breadth and rigor.

The study involved 68 elementary schools. Data were collected from each school for three sets of variables: school inputs, school social structure, and school social climate. School inputs included the socioeconomic status of students, school size, number of trained teachers per 1,000 pupils, and the like. The school social structure was defined as teacher satisfaction with the school, parental involvement in the school, and the extent to which teaching practices could be characterized as "open." School social climate was measured via 14 variables that were subdivided into student-level climate variables (e.g., sense of academic futility among pupils, appreciation and expectations pupils had for education), teacher-level climate variables (e.g., expectations about student graduation, inclination toward improving student achievement), and administrator-level climate variables (e.g., focus on academic achievement, high expectations for student achievement). Dependent variables included average achievement per school in reading and mathematics, average student self-concept, and average student self-confidence. The data were analyzed by regressing the dependent variables on the independent variables entered into the equation in a step-wise progression. Results indicated that

> when entered into the multiple regression first, the combined input set explains about 75 percent of the variance in mean school achievement, the social structures set explains 41 percent and the climate variables explain 72 percent in the representative state sample. (Brookover et al., 1979, p. 54)

In short, the three categories of variables — inputs, structure, and climate — were found to be highly related, making it difficult to determine the pattern of causality in terms of outcomes. Although the three categories of variables considered as a set accounted for a sizeable amount of variance in school-level achievement, eight percent (8%) was unique to inputs, only six percent (6%) was unique to climate, and four percent (4%) was unique to structure, again indicating a great deal of overlap between the effects of the input, structure, and climate variables. It is probably safe to say, however, that the Brookover et al. study (1978, 1979) established school climate as a central feature of effective schools. One limiting characteristic of the study was that the school was the unit of analysis, as was the case with the Rutter study. Consequently, within-school variance due to differences between individual students was not analyzed.

OUTLIER STUDIES

A significant percentage of the school effectiveness studies might loosely be referred to as outlier studies (Scheerens & Bosker, 1997). The general methodology employed in these studies was to identify those schools that are "outliers" in terms of the expected achievement of their students based on background variables (e.g., SES). Specifically, when using an outlier approach, student achievement is regressed onto various background variables and a linear, multi-variable regression equation established. Predicted achievement scores are then computed for each student and aggregated for each school. If a school's average observed achievement is *greater than* its average predicted achievement, it is considered a "positive outlier." If a school's average observed achievement is *less than* its average predicted achievement, it is considered a "negative outlier."

Purkey and Smith (1982, 1983) summarize the findings of the major outlier studies conducted up to the early 1980s, at which time the use of the outlier methodology was sharply curtailed. The studies that are the focus of their review include a study conducted by the New York State Education Department (1974a, 1974b, 1976), a study conducted by the Maryland State Department of Education (Austin, 1978, 1979, 1981), Lezotte, Edmonds, and Ratner's study (1974) of elementary schools in Detroit, Brookover and Schneider's (1975) study of elementary schools in Michigan, and Spartz's (1977) study of schools in Delaware. Despite the use of a common methodology (i.e., outliers) and a common level of schooling (i.e., elementary schools), results varied widely. For example, two of the three New York studies found that methods of reading instruction varied from high-achieving to low-achieving schools; however, one of the three studies reported no difference in instruction. Instructional leadership was one of the characteristics of effective schools identified in the Maryland study, but Spartz noted that a focus on effective administrative activities (e.g., meetings) was more critical than administrative leadership, per se. Finally, where Spartz identified seven general variables associated with high achieving schools, Brookover and Schneider identified six.

The reason for the discrepant findings in the studies is discussed in depth by Purkey and Smith (1982, 1983) and more recently by Scheerens (Scheerens, 1992; Scheerens & Bosker, 1997). Some of these shortcomings are due to the conventions of outlier methodology. They include small samples, weaknesses in the way outliers are identified owing to the fact that effects of important background characteristics are not accounted for, and regression toward the mean given that both sets of data points represent extremes. In spite of these criticisms, Scheerens and Bosker note that the following characteristics of effective schools can be inferred from the outlier research: (1) good discipline, (2) teachers' high expectations regarding student achievement, and (3) effective leadership by the school administrator.

CASE STUDIES

Another group of studies in the school effectiveness movement might be loosely referred to as case studies. In these studies, a small set of schools was studied in depth. These schools were typically organized into groups based on outcome measures — high-achieving schools versus low-achieving schools. The characteristics of schools in a group were then studied via ethnographic and/or survey techniques.

To illustrate, consider the case study by Brookover and Lezotte (1979) involving eight schools, which was a follow-up to an earlier study (Brookover et al., 1978, 1979). Brookover and Lezotte's case study focused on eight elementary schools. Five schools were defined as high need — less than 50 percent of the 4th-grade students tested attained 75 percent of the objectives on the Michigan statewide test. Three schools were defined as low need — 50 percent or more of the 4th-grade students tested attained 75 percent or more of the objectives on the statewide test. Of the low-need schools, one was defined as *improving* — it showed an increase of five percent or more in the percentage of students attaining at least 75 percent of the objectives and a simultaneous decrease of five percent or more in the percentage attaining less than 25 percent of the objectives. Two of the low-need schools were defined as *declining* — they showed a decrease of five percent or more in the percentage of students attaining at least 75 percent of the objectives and a simultaneous increase of five percent or more in the percentage of students attaining less than 25 percent of the objectives. Of the high-need schools, all five were classified as improving. A team of field researchers was sent to each site where the researchers administered questionnaires and interviewed staff members over a three- to four-day period. From this qualitative data, generalizations were constructed about the defining characteristics of effective schools. These included (1) high expectations for student achievement, (2) school policies that focus on academic achievement, (3) clear academic goals, and (4) a strong focus on basic skills.

The results of some of the more well-known case studies are reported in Table 2.2. As this table shows, these case studies had fairly homogeneous findings. The most frequently cited characteristic of effective schools, as reported in Table 2.2, is high expectations; the least frequently cited is effective staff development. All other factors were equally emphasized in the case study research.

Although it cannot be said that the case study literature led to any new insights into the characteristics of effective schools, it did help solidify the importance of the five correlates. Specifically, each variable listed in Table 2.2, with the exception of staff development, can be considered synonymous with one of the five correlates or a subcomponent of one of the five correlates. For example, "orderly climate" and "cooperative atmosphere" are analogous to "orderly atmosphere conducive to learning," and "high expectations" and "focus on basic skills" are another way of saying "high expectations for student achievement."

Table 2.2 Summary of Case Study Results

VARIABLE	STUDY			
	Weber (1971) (n = 4)[a]	**Venezky & Winfield (1979)** (n = 2)[a]	**Glenn (1981)** (n = 4)[a]	**Brookover & Lezotte (1979)** (n = 8)[a]
Strong Leadership	X		X	
Orderly Climate	X		X	
High Expectations	X	X	X	X
Frequent Evaluation	X		X	
Achievement-Oriented Policy		X		X
Cooperative Atmosphere		X	X	
Clear Academic Goals		X		X
Focus on Basic Skills		X		X
Effective Staff Development		X		

[a] Number of schools studied

IMPLEMENTATION STUDIES

Based on the assumption that the variables identified in the school effectiveness movement have a causal relationship with student achievement, a number of implementation studies were undertaken. Where all the other studies cited in this chapter were descriptive in nature, implementation studies employed interventions. In other words, an attempt was made to change school-level behavior on one or more of the factors considered important to effective schooling.

To illustrate, Milwaukee's Project RISE (McCormack-Larkin & Kritek, 1983) began in March of 1979 when the school board presented a mandate to district administrators to improve achievement in 18 elementary schools and 2 middle schools that historically had low scores on achievement tests. Project RISE was based on the assumption that the manipulation of eight critical factors can improve student achievement: (a) a shared belief that all students can learn and schools can be instrumental in that learning, (b) an explicit mission of improving student achievement, (c) high levels of professional collegiality among staff, (d) students' sense of acceptance by the school, (e) identification of grade-level objectives, (f) an accelerated program for students' achieving below grade level, (g) effective use of instructional time, and (h) a well-structured course of studies.

After three years, Project RISE schools had shown moderate increases in student achievement, particularly in mathematics. Perhaps most noteworthy about these modest gains is that they were achieved with no new staff, no new materials, and a only small amount of additional money. This, in fact, seems to be the general pattern of results for efforts to implement research from the school effectiveness movement. Specifically, the implementation studies generally indicate that focusing on the five correlates or derivatives of them produces modest gains in achievement without an expenditure of exceptional resources. (See Good and Brophy, 1986, for a discussion of efforts to implement the primary findings from the school effectiveness movement.)

CONCLUSIONS

As a whole, the school effectiveness movement produced fairly consistent findings regarding the characteristics of high-performing schools. With some variation, five general features appear to characterize effective schools as identified by a variety of methodologies, most of which focus on identifying schools where students perform better than expected based on student SES. Those five factors or five correlates as commonly referred to include (1) strong leadership, (2) high expectations for students, (3) an orderly atmosphere, (4) an emphasis on basic skills, and (5) effective monitoring of student achievement.

Chapter 3
SOME CLASSIC SYNTHESIS STUDIES

Chapter 2 discussed the research of the 1970s and early 1980s that is commonly considered to be part of the school effectiveness movement. In this chapter, studies are considered that are not part of the movement as defined in Chapter 2. Although these studies, like those from the school effectiveness movement, had as their basic purpose to articulate the defining characteristics of effective schools, many of them went beyond school characteristics to study teacher-level variables and those student-level variables that influence student achievement. In general, these studies were highly synthetic in nature in that they summarized the findings from a number of studies. In addition, many of these studies employed meta-analytic techniques as the primary data analysis strategy, providing average effect sizes (usually stated in terms of *ESd* or *r*) as the indication of the strength of the relationship between a given variable and student achievement. This chapter is organized in loose chronological order by individuals or groups of individuals who were the principal investigators for these synthetic efforts. It is safe to say that the works of these individuals and groups of individuals have come to be known as seminal studies not formally associated with the school effectiveness movement.

BLOOM

In 1984, Bloom published two articles (1984a, 1984b) that demonstrated to educators, probably for the first time, the utility of using *ESd* (the standardized mean difference) as a metric for gauging the utility of various instructional interventions. The more technical of the two articles was entitled *The 2 Sigma Problem: The Search for Methods of Instructions as Effective as One-to-One Tutoring* (1984b). The basic premise of the article was that using the most effective instructional strategies can produce achievement gains as large as those produced by one-on-one tutoring. Specifically, based on studies conducted by two of his graduate students — Anania (1982, 1983) and Burke (1984) — Bloom (1984b) concluded that tutoring has an effect size (*ESd*) of 2.00 (two sigmas) when compared with group instruction:

> It was typically found that the average student under tutoring was about two standard deviations above the average of the control class (the average tutored student was above 98% of the students in the control class). (p. 4)

Inasmuch as it is a practical impossibility to assign a tutor to every student, Bloom sought to identify "alterable educational variables" (p. 5) that would approximate the two sigma achievement effect sizes obtained by tutoring. Alterable educational variables were defined as those factors that could be reasonably influenced by teacher behavior or by resources provided by the school or district.

Bloom explicitly noted the utility of meta-analysis in the search for these variables: "Within the last three years, this search has been aided by the rapid growth of the meta-analysis literature" (p. 5). Bloom identified a number of variables that, when combined, could potentially produce a two-sigma effect. These variables were adapted from a study reported by Walberg in 1984 (discussed in the next section). They included specific instructional techniques such as reinforcement, feedback, and cooperative learning, and more general variables such as teacher expectancy. Bloom (1984b) also warned against assuming that effect sizes for different variables are additive:

> In our attempt to solve the 2 sigma problems, we assume that two or three alterable variables must be used that *together* contribute more to learning than any one of them. . . . So far, we have *not* found any two variable combinations that have exceeded the 2 sigma effect. Thus, some of our present research reaches the 2 sigma effect, but does not go beyond it. (p. 6)

Both of Bloom's 1984 articles (1984a, 1984b) also extolled the powerful effects of mastery learning (ML). For example, Bloom (1984b) wrote:

> Because of more than 15 years of experience with ML at different levels of education and in different countries, we have come to rely on ML as one of the possible variables to be combined with selected other variables. ML (the feedback-corrective process) under good conditions yields approximately a 1 sigma effect size. (p. 6)

Although Bloom's work and that of his colleagues is sometimes thought of in the narrow context only of mastery learning, in fact Bloom was probably the first researcher to demonstrate, via the use of the *ESd* index, the powerful influence that effective instruction can have on student achievement.

WALBERG

It is probably safe to say that Walberg has been one of the most prominent figures in the last 20 years relative to attempts to identify those factors that most strongly influence school learning. Most of his writings make explicit reference to his "productivity model," which was first articulated in 1980 in a publication entitled *A Psychological Theory of Educational Productivity*. In that article, Walberg argued that achievement in school can be described as a function of seven factors:

1. student ability (*Abl*)
2. motivational factors (*Mot*)
3. quality of instruction (*Qal*)
4. quantity of instruction(*Qan*)
5. classroom variables (*Clas*)
6. home environment (*Home*)
7. age or mental development (*Age*)

Walberg further argued that the most appropriate mathematical model to describe the extent to which these factors predict achievement is the Cobb-Douglas (1928) function borrowed from economics, as opposed to a more traditional linear regression model. The general form of the Cobb-Douglas function is $O = aK^bL^c$, where O is output or productivity, a is a constant, K is capital, L is labor, and b and c are exponents. When Walberg applied this function to his seven factors, the following equation resulted:

$$\text{Achievement} = a \text{ x}(Abl)^b\text{x}(Mot)^c\text{x}(Qal)^d\text{x}(Qan)^f\text{x}(Cls)^g\text{x}(Hom)^h\text{x}(Age)^I$$

Walberg (1980) detailed the many advantages of the Cobb-Douglas function, two of which are

- increasing the productivity or effectiveness of one factor while keeping the others constant produces diminishing returns
- a zero value for any factor will return a product of zero. (pp. 14–15)

These aspects of the Cobb-Douglas function had great intuitive appeal for Walberg in the context of predicting student achievement. For example, it makes intuitive sense that increasing the quantity of instruction without increasing any of the other six factors in Walberg's model will have diminishing returns on achievement over time. Similarly, a value of zero for motivational factors, for example, will produce zero achievement regardless of the values assigned to the other six factors.

In a 1984 article entitled "Improving the Productivity of America's Schools," Walberg expanded on his productivity model.[1] In this later work, Walberg identified nine factors organized into three general categories:

A. Student Aptitude
   1. Ability or prior achievement
   2. Development as indexed by age or stage of maturation
   3. Motivation or self-concept as described by personality tests or the student's willingness to persevere intensively on learning tasks

B. Instruction
   1. The amount of time students are engaged
   2. The quality of instruction

C. Environment
   1. The home
   2. The classroom social groups
   3. The peer groups outside of school
   4. Use of out-of-school time (specifically, the amount of leisure time television viewing)

In defense of the model, Walberg (1984) reported that "about 3,000 studies suggest that these factors are the chief influences on cognitive, affective, and behavioral learning" (p. 22). Although Walberg reported average effect sizes for a variety of variables in each of the nine categories, he mixed different types of effect sizes (i.e., correlations versus standardized mean differences) without specifying which metric was being used, making it difficult, if not impossible, to ascertain the relative impact of the various factors. Nevertheless, Walberg's productivity model has been in the forefront of many discussions about variables that influence student achievement, particularly in the last decade.

---

[1]This is the article from which Bloom (1984a, 1984b) derived his list of alterable variables.

Fraser, Walberg, Welch, and Hattie

In 1987, an issue of the *International Journal of Educational Research* was devoted to a summary of the research on school- and classroom-level variables affecting achievement. The volume contained six chapters written (without designating chapter authorship) by Fraser, Walberg, Welch, and Hattie. The overall title of the volume was "Synthesis of Educational Productivity Research," signaling the strong influence of Walberg's productivity model. Indeed, the first chapter of the volume addressed the need for a major review of the literature and the utility of using meta-analysis as the synthetic technique with which to review the literature. It then specified Walberg's (1984) nine-factor productivity model as that which would be used to organize the findings presented in the volume. Three separate sets of findings were reported.

The first set of findings utilized Walberg's productivity model to synthesize the results of 2,575 individual studies. This synthesis was identical to Walberg's 1984 article, which was used by Bloom in his two 1984 articles. As was the case with the 1984 Walberg article, Fraser et al. utilized reporting conventions that made it difficult to interpret the findings. The overall conclusion of this first set of findings was that "the first five essential factors in the educational productivity model (ability, development, motivation, quantity of instruction, quality of instruction) appear to substitute, compensate, or trade off for one another at diminishing rates of return" (p. 163).

The centerpiece of the journal issue was a section entitled "Identifying the Salient Facets of a Model of Student Learning: A Synthesis of Meta-Analyses." It synthesized the results of 134 meta-analyses, which were based on 7,827 studies and 22,155 correlations. An estimated 5–15 million students in kindergarten through college were involved in these studies as subjects. Seven factors that are clearly related, but not identical, to the nine factors in Walberg's productivity model were used to organize the findings: (1) school factors, (2) social factors, (3) instructor factors, (4) instructional factors, (5) pupil factors, (6) methods of instruction, and (7) learning strategies. The average correlation with achievement across all seven factors was .20 (*ESd* = .41). The correlations and effect size (*ESd*) for each of these seven factors are reported in Table 3.1.

Unlike the first set of findings reported in the Fraser et al. study, those summarized in Table 3.1 provided specific information about the number of studies involved, the specific studies that were used, and the variability and central tendency of the findings for different variables. In fact, the results reported in Table 3.1 are still considered by many to be the most comprehensive review of research in terms of the number of studies involved.

The third set of findings reported by Fraser et al. was specific to the science achievement of 17-, 13-, and 9-year-olds in the United States in 1981–82. The study incorporated data from studies involving 1,955 seventeen-year-olds, 2,025 thirteen-year-olds, and 1,960 nine-year-olds. Loosely speaking, seven of Walberg's nine factors were used to organize the data. The correlations and effect sizes for each of the three age groups for each factor are reported in Table 3.2.

Table 3.1
Summaries of the Relationships of Factors to Achievement

Factor	No. of Meta-Analyses	No. of Studies	No. of Relationships	Average *r*	Average *ESd*
1. School	16	781	3,313	.12	.25
2. Social	4	153	1,124	.19	.39
3. Instructor	9	329	1,097	.21	.44
4. Instruction	31	1,854	5,710	.22	.47
5. Pupil	25	1,455	3,776	.24	.47
6. Methods of Instruction	37	2,541	6,352	.14	.29
7. Learning Strategies	12	714	783	.28	.61
Overall	134	7,827	22,155	.20	.41

Note: Adapted from "Syntheses of Educational Productivity Research," by B. J. Fraser, H. J. Walberg, W. A. Welch, and J. A. Hattie, 1987, *International Journal of Educational Research 11*(2) [special issue], p. 207.

r is the Pearson product-moment correlation coefficient; *ESd* is Cohen's effect size *d*.

Table 3.2
Science Achievement
Correlation and Effect Size by Productivity Factor for Three Age Levels

Factor	17-year-olds		13-year-olds		9-year-olds	
	r	***ESd***	***r***	***ESd***	***r***	***ESd***
Ability	.42	.926	.30	.629	.48	1.094
Motivation	.27	.561	.23	.473	.25	.516
Quality of Instruction	.09	.181	.09	.181	.01	.020
Quantity of Instruction	.31	.652	.23	.473	0.00	0.00
Class Environment	.23	.473	.25	.516	.14	.283
Home Environment	.27	.561	.18	.366	.16	.324
Television	–.16	-.324	–.09	-.181	–.10	-.201

Note: Adapted from "Syntheses of Educational Productivity Research," by B. J. Fraser, H. J. Walberg, W. A. Welch, and J. A. Hattie, 1987, *International Journal of Educational Research 11*(2) [special issue], p. 220.
r is the Pearson product-moment correlation coefficient; *ESd* stands for Cohen's effect size *d*.

It is instructive to note that the seven factors used as the organizational framework in Table 3.2 are defined quite differently from those in Table 3.1. For example, in Table 3.2, quality of instruction is defined as the total budget allocated for science instruction in a school; in Table 3.1, quality of instruction, a sub-factor of "Instruction," addresses specific types of instructional techniques. These differences in definitions most likely account for the differences in findings reported by Fraser et al. For example, Table 3.2 reports correlations of .09 and .01 for quality of instruction and student achievement; however, relative to the science achievement findings, the researchers reported an average correlation of .47 for quality of instruction and student achievement (see Fraser et al., 1987).

Although the Fraser et al (1987) monograph reported multiple findings, it concluded with an explicit validation of Walberg's productivity model: "Overall, then, the work reported throughout the monograph provides much support for most of the factors in the productivity model in influencing learning" (p. 230). Although this conclusion probably goes beyond the data reported, the Fraser et al. report was a milestone in the research on those factors that influence student achievement. Specifically, its review of 134 meta-analyses (see Table 3.1) provided some compelling evidence that the research literature considered as a whole supports the hypothesis that schools can make a difference in student achievement. This conclusion was made even more explicit by one of the volume's authors, John Hattie.

HATTIE

Hattie was one of the coauthors of the Fraser et al. special issue of *The International Journal of Educational Research*. Specifically, Hattie was the primary author of the volume's section entitled "Identifying the Salient Facets of a Model of Student Learning: A Synthesis of Meta-Analyses." As described above, this section synthesized the results of 134 meta-analyses and was considered the centerpiece of the volume.

In 1992, Hattie republished these findings under his own name in an article entitled "Measuring the Effects of Schooling." However, in this later publication, he more strongly emphasized a number of salient findings from the synthesis of the 134 meta-analyses. First, he emphasized the practical significance of the average effect size across the seven factors used to categorize the data (i.e., school, social, instructor, instruction, pupil, methods of instruction, and learning strategies) from the 7,827 studies and 22,155 effect sizes. Hattie explained:

> Most innovations that are introduced in schools improve achievement by about .4 standard deviations. This is the benchmark figure and provides a standard from which to judge effects — a comparison based on typical, real-world effects rather than based on the strongest cause possible, or with the weakest cause imaginable. At a minimum, this continuum provides a method for measuring the effects of schooling. (p. 7)

Further, Hattie (1992) decomposed this average effect size into useful components. Specifically, based on Johnson and Zwick's (1990) analysis of data from the National Assessment of Educational Progress, Hattie reasoned that one could expect a gain in student achievement of .24 standard deviations in a school where no innovations were used — in nontechnical terms, one might say that a "regular" school produces an effect size (*ESd*) of .24. Using the research of Cahen and Davis

(1977), Hattie further reasoned that about 42 percent of the effect size of .24 is due simply to student maturation. Thus, one could expect a regular school to produce an achievement gain of .14 standard deviations above and beyond that from maturation (which is .10). Finally, Hattie reasoned that the innovations identified in his meta-analyses increased achievement by .16 standard deviations above and beyond maturation and regular schooling. Hattie was perhaps the first to provide this perspective on the effects of maturation versus regular schooling and versus "innovative" schooling.

Hattie (1992) also articulated three major conclusions that could be drawn from his meta-analysis. First, he noted that one theme underlying the findings was that a "constant and deliberate attempt to improve the quality of learning on behalf of the system . . . typically relates to improved achievement" (p. 8). Second, Hattie explained that "the most powerful, single moderator that enhances achievement is feedback. The simplest prescription for improving education must be 'dollops of feedback'" (p. 9). Third, Hattie noted that strategies that focus on individualizing instruction do not have great success: "Most innovations that attempt to individualize instruction are not noted by success" (p. 9). He further explained that this is particularly disturbing especially in light of Rosenshine's (1979) research indicating that students spend about 60 percent of their time working alone.

In 1996, Hattie, Biggs, and Purdie published the results of a second meta-analysis that synthesized the findings from 51 different studies of instructional practices involving 270 effect sizes. The primary, independent variable and, hence, organizer for the meta-analysis was a taxonomy developed by Biggs and Collis (1982). The taxonomy includes four levels of cognitive tasks:

- Level 1: Unistructional Tasks: Skills taught in a step-by-step fashion.
- Level 2: Multistructional Tasks: Skills taught that involve multiple strategies, but with little or no emphasis on the metacognitive aspects of the processing.
- Level 3: Relational Tasks: Multiple skills taught with an emphasis on the metacognitive aspects of the processing.
- Level 4: Extended Abstract: Multiple skills taught with an emphasis on application to new domains.

The results of this meta-analysis are summarized in Table 3.3. One obvious inconsistency in the findings reported in Table 3.3 is the lack of a taxonomic-like pattern in the effect sizes. Specifically, Hattie et al. (1996) hypothesized that the extended abstract tasks would produce greater learning (i.e., a higher effect size) than the relational tasks, which would produce greater learning than the multi-instructional tasks, which would produce greater learning than the uninstructional task if the taxonomy were valid. But this is not what they found. The researchers explain these unpredicted findings as a function of the types of dependent measures that were used as opposed to possible problems with the classification system.

Taken together, Hattie's synthetic efforts contributed significantly to the knowledge base about schooling. His re-analysis of the Fraser et al. (1987) data provided a new perspective on the results. The results of the Hattie et al. (1996) meta-analysis also added new insights to the growing research base on instructional practices.

Table 3.3
Summary of Findings From Hattie et al. 1996 Meta-Analysis

Nature of Intervention	*N*	*ESd*
Unistructional	29	.84
Multistructional	16	.45
Relational	34	.22
Extended Abstract	40	.69

Note: Constructed from "Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis," by J. Hattie, J. Biggs, and N. Purdie, 1996, *Review of Educational Research, 66*(2), 99–136.

N is the number of studies; *ESd* stands for Cohen's effect size *d*.

WANG, HAERTEL, AND WALBERG

Perhaps the most robust attempt to synthesize a variety of research and theoretical findings on the salient variables affecting school learning was conducted by Wang, Haertel, and Walberg (1993). The final report on this effort was in an article entitled "Toward a Knowledge Base for School Learning." This publication became the basis for a number of other publications (e.g., Wang, Reynolds, & Walberg, 1994; Wang, Haertel, & Walberg, 1995). The 1993 Wang et al. article combined the results of three previous studies. Although not the first chronologically, the conceptual centerpiece of the three studies was reported by Wang, Haertel, and Walberg (1990). It involved a comprehensive review of the narrative literature on school learning. The review addressed literature in both general and special education including relevant chapters in the American Educational Research Association's *Handbook of Research on Teaching* (Wittrock, 1986), the four-volume *Handbook of Special Education: Research and Practice* (Wang, Reynolds, & Walberg, 1987–1991), *Designs for Compensatory Education* (Williams, Richmond, & Mason, 1986), and the various annual review series that are reported in education, special education, psychology, and sociology. In total, the synthesis covered 86 chapters from annual reviews, 44 handbook chapters, 20 government and commissioned reports, 18 book chapters, and 11 journal articles.

The review encompassed 3,700 references and produced 228 variables identified as potentially important to school learning. A rating on a 3-point scale was assigned by Wang, Haertel, and Walberg to each citation indicating the strength of the relationship between the variable and school learning. The 228 variables were then collapsed into 30 categories, which were grouped into seven broad domains: (1) state and district variables, (2) out-of-school contextual variables, (3) school-level variables, (4) student variables, (5) program design variables, (6) classroom instruction, and (7) climate variables.

The second study in the triad was reported by Reynolds, Wang, and Walberg (1992). The study surveyed 134 education research experts who were first authors of the major annual reviews and handbook chapters, book chapters, government documents, and journal review articles used in the Wang et al. (1990) study. These experts were surveyed and asked to rate the 228 variables on a 4-

point Likert scale indicating the influence of each of the 228 variables on student learning. The scale ranged from 3, indicating strong influence on learning, to 2, indicating moderate influence, to 1, indicating little or no influence, to 0, indicating uncertain influence on learning. Forty-six percent (46%) of the experts responded to the survey. Mean scores were calculated for each of the 228 variables. These mean ratings were then used to compute the mean ratings for the 30 categories and seven domains formulated in the Wang et al. (1990) study.

The third study in the triad was the six-chapter issue of the *International Journal of Educational Research* by Fraser and his colleagues (1987). As described previously, this study synthesized the results of 134 meta-analyses. The Wang et al. (1993) study utilized 130 of the 134 meta-analyses along with the results from six meta-analyses not addressed by Fraser et al. (1987), resulting in a data base of 136 meta-analyses. Wang et al. (1993) determined that the 136 meta-analyses addressed only 23 of the 30 categories identified in the Wang et al. (1990) and the Reynolds et al. (1990) studies. A weighted mean correlation was computed for each of these 23 variables.

To combine the results from the three studies, the mean ratings for the Wang et al. (1990) content analyses, the mean ratings from the education experts survey by Reynolds, Wang, and Walberg (1992), and the weighted mean correlations from the Fraser, Walberg, Welch, and Hattie (1987) study were transformed into *Z* scores. The *Z* scores were then transformed into *T* scores (i.e., scaled scores) with a mean of 50 and a standard deviation of 10.

The 30 variables were then organized into six categories referred to as the six "theoretical constructs" by Wang et al. (1993): (1) student characteristics, (2) classroom practices, (3) home and community education context, (4) design and delivery of curriculum and instruction, (5) school demographics, culture, climate, policies and practices, and (6) state and district governance and organizations. Average *T* scores were calculated for each of these six theoretical constructs. These are listed in Table 3.4.

Table 3.4
***T* Scores for Wang et al.'s (1993) Theoretical Constructs**

Theoretical Construct	Average *T* score
Student characteristics	54.7
Classroom practices	53.3
Home and community educational contexts	51.4
Design and delivery of curriculum and instruction	47.3
School demographics, culture, climate, policies & practices	45.1
State and district governance	35.0

Note: See "Toward a Knowledge Base for School Learning," by M. C. Wang, G. D. Haertel, and H. J. Walberg, 1993, *Review of Educational Research, 63*(3), p. 270.

Average *T* scores also were computed for the 30 variables that made up the six theoretical constructs. The top five variables in descending order of importance as defined by their *T*-score values were

- classroom management
- student use of metacognitive strategies
- student use of cognitive strategies
- home environment and parental support
- student and teacher social interactions

The five variables with the weakest relationship to school learning as defined by their *T*-score values were

- program demographics
- school demographics
- state and district policies
- school policy and organization
- district demographics

Based on the composite findings, Wang, Haertel, and Walberg concluded that "proximal" variables — those closest to students — have a stronger impact on school learning than do "distal" variables — those somewhat removed from students. Given the breadth of the effort, the Wang et al. (1993) study is frequently cited in the research literature as a state-of-the-art commentary on the variables that affect student achievement.

LIPSEY AND WILSON

In 1993, psychologists Lipsey and Wilson conducted a meta-analysis of 302 studies that cut across both education and psychotherapy. Their purpose was to provide an overview of the effects of various categories of educational and psychological interventions on a variety of outcomes. The results for the various subcategories in education are reported in Table 3.5.

The mean effect size (*ESd*) across all studies (education and psychology) was .50 (SD = .29, N = 302 studies, 16,902 effect sizes). It is interesting to note that this average effect size is relatively close to that reported of .40 by Hattie in 1992. The relatively large average effect size was considered so striking by Lipsey and Wilson that it led them to comment: "Indeed, the effect size distribution is so overwhelmingly positive that it hardly seems plausible that it presents a valid picture of the efficacy of treatment per se" (p. 1192).

Perhaps the biggest contribution of the Lipsey and Wilson meta-analysis was its detailed examination of a variety of moderator variables commonly addressed in meta-analyses. Specifically, Lipsey and Wilson analyzed the differential effects on the interpretation of effect sizes of (1) methodological quality, (2) publication bias, and (3) small sample bias.

Table 3.5
Findings from Education Studies

Studies	*N*	*Average ESd*
1.0 General Education, K–12 and College		
1.1 Computer aided/based instruction	622	0.362
1.2 Programmed or individualized instruction	724	0.296
1.3 Audio and visual based instruction	215	0.339
1.4 Cooperative task structures	414	0.629
1.5 Student tutoring	430	0.821
1.6 Behavioral objectives, reinforcement, cues, feedback, etc.	204	0.546
1.7 Other general education	546	0.327
2.0 Classroom Organization/Environment		
2.1 Open classroom vs. traditional	295	-0.056
2.2 Class size	213	0.295
2.3 Between and within class ability grouping	224	0.119
2.4 Other classroom organization/environment	20	0.476
3.0 Feedback to Teachers	218	0.776
4.0 Test Taking		
4.1 Coaching programs for test performance	210	0.275
4.2 Test anxiety	674	0.649
4.3 Examiner	22	0.35
5.0 Specific Instructional or Content Areas		
5.1 Science and math instruction	1769	0.310
5.2 Special content other than science and math	697	0.497
5.3 Preschool and special education; developmental disabilities		
5.3.1 Early Intervention for disadvantaged or handicapped	293	0.445
5.3.2 Special education programs or classrooms	277	0.503
5.3.3 Perceptual-motor and sensory stimulation treatment	318	0.264
5.3.4 Remedial language programs and bilingual	154	0.587
5.3.5 Other special education	265	0.731
5.4 Teacher training		
5.4.1 In-service training for teachers	464	0.593
5.4.2 Practice or field experience during teacher training	85	0.184
6.0 Miscellaneous Educational Interventions	635	0.487

Note: Constructed from data in "The Efficacy of Psychological, Educational, and Behavioral Treatment," by M. W. Lipsey and D. B. Wilson, 1993, *American Psychologist, 48*(12), 1181–1209. *N* is the number of studies. *ESd* stands for Cohen's effect size *d*.

It is frequently assumed that studies that use more rigorous research designs will have lower effect sizes since they control for systematic variation not of experimental interest that might inflate effect size estimates. However, Lipsey and Wilson found that there is no difference (i.e., statistically significant differences) between effect sizes from studies rated high in methodological quality versus those rated low. Neither were there differences in effect sizes for studies that used random assignment to experimental and control groups versus those that use nonrandom assignments. However, there was a .29 differential between effect sizes that were computed from comparison of experimental versus control groups and those from one-group, pre-post test designs with the latter design having the larger effect size.

Another factor that is thought to inflate effect size estimates in the context of a meta-analysis is systematic differences between studies that are published versus those that are not published. The general assumption is that studies with statistically significant effect sizes will be published; those that do not report significant effect sizes will not. Therefore, if a meta-analysis samples only those studies that are published, the sample will be biased upwards, producing artificially high effect sizes. Lipsey and Wilson found that within their sample, published studies yielded mean effect sizes that averaged .4 *SD*s larger than unpublished studies. They noted that "it is evident, therefore, that treatment effects reported in published studies are indeed generally biased upward relative to those in unpublished studies" (p. 1195).

The third moderator variable studied by Lipsey and Wilson was sample size. It has been demonstrated conceptually that mean effect sizes based on small samples are biased upward as a statistical estimator of the population effect size means (see Hedges & Olkin, 1985). Consequently, the mean effect size in a meta-analysis that includes a high proportion of studies that use a small sample size might have a bias toward overestimation. To study this statistical phenomenon, Lipsey and Wilson compared the average effect size for studies with less than 50 subjects and those with more than 50 subjects. No significant difference was found between these two means, indicating that small sample bias was not operating in their study.

Although the Lipsey and Wilson study is not commonly cited in the research literature in education, it is a valuable addition to the research base. First, it added significantly to the mounting body of evidence that schools can make a difference. Also, it helped establish meta-analysis as a viable tool for synthesizing the research on schooling.

COTTON

Cotton's (1995) study was one of the most comprehensive of narrative reviews in that it included over 1,000 citations. Narrative reviews are much more inductive and qualitative in nature than are meta-analytic reviews. Where meta-analytic reviews rely on interpretations of mathematical averages of effect sizes computed for each study, narrative reviews rely on interpretations of the subjective conclusions from the studies that are being synthesized. In spite of the fact that narrative reviews have been shown to be subject to considerable error in their interpretation of findings (see Cooper & Rosenthal, 1980), they are still far more common than meta-analytic reviews of schooling.

Cotton's review identified variables associated with student achievement at the classroom level, the school level, and the district level. The major variables associated with these three levels are summarized in Table 3.6. For each of the variables reported in Table 3.6, Cotton listed more specific elements. For example, in the school-level variable of "leadership and school improvement," Cotton lists the following subcomponents:

1. Leaders undertake school restructuring efforts as needed to attain agreed-upon goals for students. . . .
2. Strong leadership guides the instructional program. . . .
3. Administrators and other leaders continually strive to improve instructional effectiveness. (pp. 28–29)

Table 3.6
Three Levels of Variables in Cotton's Review

Classroom-level variables:

- planning and setting goals
- classroom management and organization
- instruction
- teacher-student interactions
- equity
- assessment

School-level variables:

- planning and learning goals
- school management and organization
- leadership and school improvement
- administrator-teacher-student interactions
- equity
- assessment
- special programs
- parent and community involvement

District-level variables:

- leadership and planning
- curriculum
- district-school interaction
- assessment

Note: See *Effective Schooling Practices: A Research Synthesis. 1995 Update*, by K. Cotton, 1995, School Improvement Research Series. Portland, OR: Northwest Regional Educational Laboratory.

Within these subcomponents, Cotton identifies even more specific characteristics. For example, the following characteristics are listed under the first subcomponent:

Administrators and other leaders

- Review school operations in light of agreed-upon goals for student performance.
- Work with school-based management team members to identify any needed changes (in organization, curriculum, instruction, scheduling, etc.) to support

attainment of goals for students.
- Identify kinds of staff development needed to enable school leaders and other personnel to bring about desired changes.
- Study restructuring efforts conducted elsewhere for ideas and approaches to use or adapt.
- Consider school contextual factors when undertaking restructuring efforts factors such as availability of resources, nature of incentive and disincentives, linkages within the school, school goals and priorities, factions and stresses among the staff, current instructional practices, and legacy of previous innovations. (p. 28)

Cotton's review is certainly impressive in its breadth. One criticism of the review, however, is that it does little to synthesize the research findings into manageable units. To illustrate, at the classroom level over 160 elements are listed, at the school level over 220 elements are listed, and at the district level over 50 elements are listed. Such a daunting list does little for district-, school-, or classroom-level educators seeking to make meaningful change. Another shortcoming of the Cotton review is that it provides no explanation of how categories are formed and how components and subcomponents in each category are identified. Additionally, Cotton offers no discussion of the frequency with which the various elements she identifies are cited in the 1,000-plus references that accompany the review.

SCHEERENS AND BOSKER

One of the most quantitatively sophisticated reviews of the research literature on factors influencing student achievement is that conducted by Scheerens and Bosker (see Scheerens & Bosker,1997; Scheerens, 1992; Bosker, 1992; Bosker & Witziers, 1995, 1996). The overall mathematical model used to organize the research was a hierarchical linear model (HLM).[2] The centerpiece of Scheerens and Bosker's work was a meta-analysis of an international literature base of the effects of nine factors on student achievement:

1. *Cooperation:* The extent to which staff members in a school supported one another, sharing resources, ideas, and problem solutions.
2. *School Climate:* The extent to which the school has an achievement-oriented culture and maintains order in a positive manner.
3. *Monitoring:* The extent to which the school seeks out and uses feedback relative to whether it is accomplishing its academic goals.
4. *Content Coverage:* The extent to which the school monitors the coverage of the identified curriculum.
5. *Homework:* The extent to which the school articulates and implements a homework policy.
6. *Time:* The amount of time a school allots for instruction.
7. *Parental Involvement:* The extent to which parents are involved in the functions of the school.

[2] The specifics of HLM and how it might be used are discussed in some depth in Chapter 7 and, therefore, will not be addressed here.

8. *Pressure to Achieve:* The extent to which the school communicates a strong message that academic achievement is a primary goal.
9. *Leadership:* The extent to which the school has strong leadership relative to the goal of academic achievement.

The specific effect sizes associated with these factors will be discussed in Chapter 4 in some depth and, thus, are not reported here. Suffice it to say that the Scheerens and Bosker study provides the most rigorous analysis of the research on these variables to date. In addition to thoroughly discussing the nine factors just summarized, Scheerens and Bosker summarized the research from qualitative reviews, international analyses, and research syntheses on a number of school-level factors that affect achievement. This review is presented in Table 3.7. The synthesis reported in Table 3.7 is unique in that it offers a comparison of qualitative syntheses with quantitative syntheses. Of particular note is the pattern of support across all three literature bases for academic pressure to achieve, parental involvement, orderly climate, and opportunity to learn.

CREEMERS

Using a narrative approach, Creemers (1994) synthesized much of the same research that Scheerens and Bosker synthesized. Creemers used the model shown in Figure 3.2 as the basic organizing scheme for his synthesis. He refers to this as the basic model of "educational effectiveness." Within this general model, Creemers focused attention on quality of instruction. He offered the synthesis of research reprinted in Table 3.8.

Creemers' coding of instructional strategies in terms of strong empirical evidence, moderately empirical evidence, and plausible empirical evidence makes for a rather straightforward interpretation of the classroom-level variables he identifies. Unfortunately, he offers little or no explanation for his codings even though some seem to fly in the face of current research and conventional wisdom. For example, in Table 3.8 cooperative learning has an overall rating of plausible only. However, a meta-analysis by Johnson et al. (1981) indicates that cooperative learning has an average effect size (*ESd*) of .73 which is considered high moderate to large (Cohen, 1988). With these problems acknowledged, it is only fair to state that Creemers' work is probably considered the most comprehensive analysis of the research on instruction to date.

Table 3.7
Summary of Evidence from Qualitative, International, And Synthetic Studies

Categories	Qualitative reviews	International analyses	Research syntheses
Resource input variables:			
Pupil-teacher ratio		–0.03	0.02
Teacher training		0.00	–0.03
Teacher experience			0.04
Teachers' salaries			–0.07[a]
Expenditure per pupil			0.20[b]
School organizational factors:			
Productive climate culture	+		
Achievement pressure for basic subjects	+	0.02	0.14
Educational leadership	+	0.04	0.05
Monitoring/evaluation	+	0.00	0.15
Cooperation/consensus	+	–0.02	0.03
Parental involvement	+	0.08	0.13
Staff development	+		
High expectations	+	0.20	
Orderly climate	+	0.04	0.11
Instructional conditions:			
Opportunity to learn	+	0.15	0.09
Time on task/homework	+	0.00/–0.01 (n.s.)	0.19/0.06
Structured teaching	+	–0.01 (n.s.)	0.11 (n.s.)
Aspects of structured teaching:			
– cooperative learning			0.27
– feedback			0.48
– reinforcement			0.58
Differentiation/adaptive instruction			0.22

Note: Reprinted from *The Foundations of Educational Effectiveness* (p. 305), by J. Scheerens and R. J. Bosker, 1997, New York: Elsevier, with the permission of Elsevier Science.
Numbers refer to correlations, the size of which might be interpreted as 0.10: small; 0.30: medium; 0.50: large (cf. Cohen, 1988).
+ indicates positive influence; n.s. indicates statistically not significant.
[a]Having assumed a standard deviation of $5,000 for teacher salary.
[b]Assuming a standard deviation of $100 for PPE.

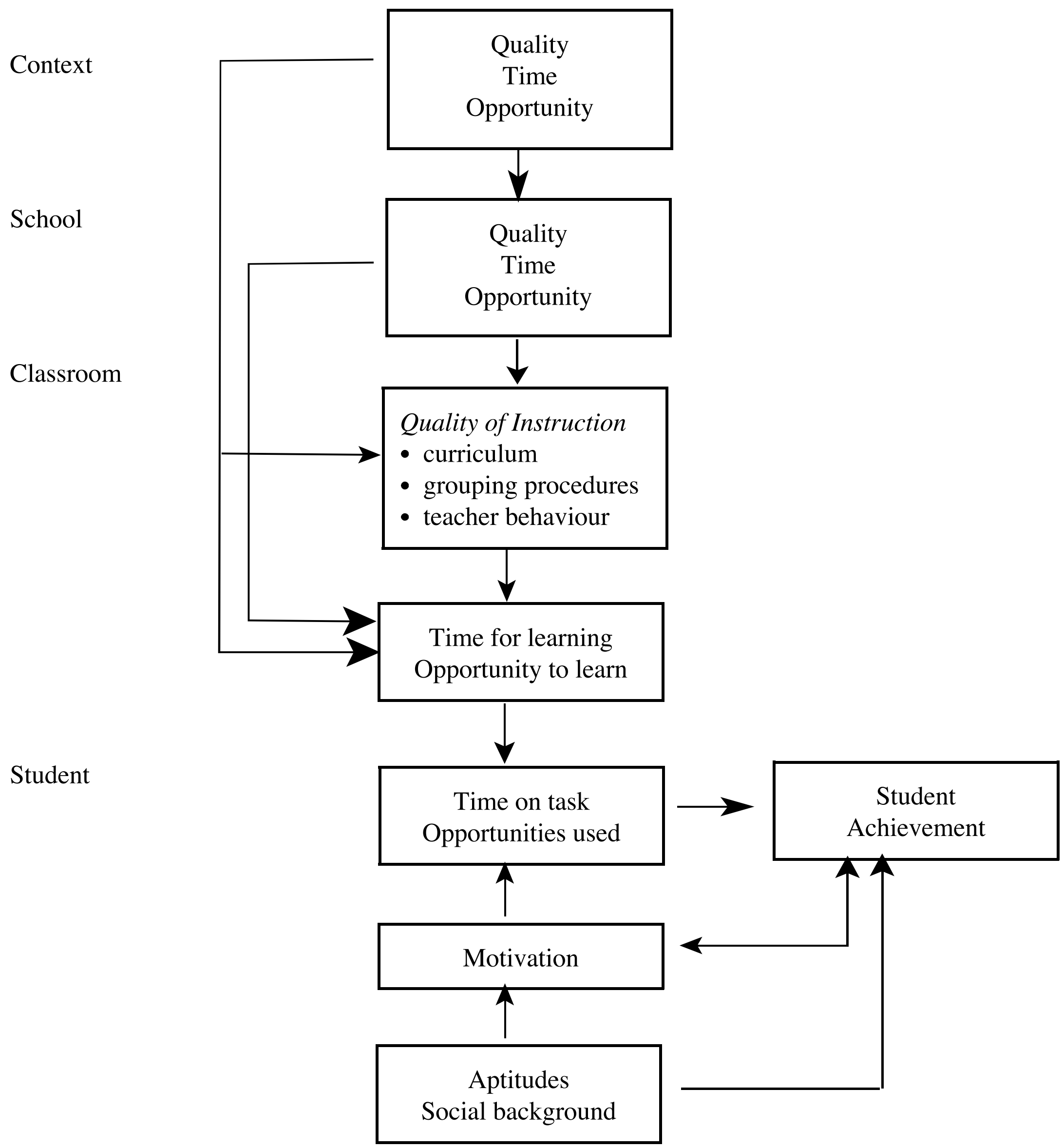

Figure 3.2. The basic model of educational effectiveness.

Note: From *The Effective Classroom* (p. 27), by B. P. M. Creemers, 1994, London: Cassell. Reprinted with permission.

Table 3.8
Overview of Empirical Evidence for the Characteristics of Effective Instruction

Characteristics	Strong empirical evidence	Moderate empirical evidence	Plausible
Curriculum		x	
Grouping procedures	x		
Teacher behaviour	x		
Curriculum			
Explicitness and ordering of goals and content	x		
Structure and clarity of content		x	
Advance organizers	x		
Evaluation	x		
Feedback	x		
Corrective instruction			x
Grouping procedures			
Mastery learning	x		
Ability grouping		x	
Cooperative learning			x
Differentiated material			x
Evaluation	x		
Feedback		x	
Corrective instruction		x	
Teacher behaviour			
Management/orderly and quiet atmosphere	x		
Homework	x		
High expectations		x	
Clear goal setting		x	
Restricted set of goals		x	
Emphasis of basic skills		x	
Emphasis on cognitive learning and transfer			x
Structuring the content		x	
Ordering of goals and content		x	
Advance organizers	x		
Prior knowledge		x	
Clarity of presentation		x	
Questioning	x		
Immediate exercise		x	
Evaluation	x		
Feedback		x	
Corrective instruction			x

Note: From *The Effective Classroom* (p. 94), by B. P. M. Creemers, 1994, London: Cassell. Reprinted with permission.

THREE CATEGORIES OF VARIABLES

From the discussion in this chapter and the preceding chapter, it should be evident that there are multiple ways to organize the research on variables that affect student achievement. However, one organizational pattern does seem to cut across a multitude of studies. Specifically, the following three categories appear to be implicit or explicit in a variety of studies: (1) school-level variables, (2) teacher-level variables, and (3) student-level variables. To illustrate, Table 3.9 summarizes the extent to which a number of popular models utilize these categories.

Table 3.9
Three Categories of Variables

Study	School Level	Teacher Level	Student Level
Elberts & Stone (1988)	I	E	E
Carroll (1963, 1989)	I	E	E
Rowe, Hill & Holmes-Smith (1993)	E	E	E
Walberg (1984)	I	E	E
Scheerens (1990)	E	E	E
Creemers (1994)	E	E	E
Scheerens & Bosker (1997)	E	E	E
Cotton (1995)	E	E	E
Wright, Horn, & Sanders (1997)	E	E	E
van der Werf (1997)	E	E	E
Goldstein (1997)	I	E	E
Raudenbush & Bryk (1988)	E	E	E
Raudenbush & Willms (1995)	E	E	E

Note: E indicates that the categories were explicitly used in the study; I indicates that the three categories were implicit.

As Table 3.9 shows, all of the 13 studies reviewed explicitly use the teacher and student levels as primary organizers for the variables affecting student achievement. In addition, 9 out of the13 explicitly use the school level as a primary organizer, and the remaining 4 use the school level implicitly as an organizer. Given the wide acceptance of these levels as organizers, they are employed in the remainder of this monograph.

Both the school effectiveness research reviewed in Chapter 2 and the quantitative and qualitative synthesis reviewed in this chapter support the hypothesis that certain identifiable variables have a significant impact on student achievement. In Part II, the variables specific to schools, teachers, and students are reviewed with an eye to their unique effects and their composition.

PART II:
RESEARCH ON SCHOOL, TEACHER, AND STUDENT EFFECTS

Chapter 4
THE SCHOOL-LEVEL EFFECT

This chapter focuses on school-level variables that influence student achievement. In effect, this chapter seeks to answer the questions, How large is the school effect? and What school-level variables comprise that effect? Raudenbush and Willms (1995) make a distinction between two types of school-level effects that are useful to this discussion. They begin with the model shown in Table 4.1.

Table 4.1
Raudenbush and Willms' Model

$Y_{ij} = u + P_{ij} + C_{ij} + S_{ij} + e_{ij}$
• Y_{ij} is the achievement of student *i* in school *j* • *u* is the grand mean for all student achievement scores • P_{ij} is the effect of school practice (e.g., policies of the school, resources of the school, instructional leadership, effectiveness of classroom practice, and so on) • C_{ij} is the contribution of the school context (i.e., the socioeconomic status of the neighborhood in which the school resides, the employment rate of the community, and so on) • S_{ij} is the influence of background variables specific to each student (e.g., student aptitude, the socioeconomic status of each student, and so on) • e_{ij} is a random error term including unmeasured sources of a particular student's achievement assumed to be statistically independent of *P*, *C*, and *S*

Note: See "The Estimation of School Effects," by S. W. Raudenbush and J. D. Willms, 1995, *Journal of Educational and Behavioral Statistics, 20*(4), 307–335.

An important feature of the model is that *P* and *C* are allowed to vary across students in a school. That is, there is no assumption that school practices or school context affect all students the same way — hence, the use of the subscripts *i* and *j* with the *P* and *C* terms in the model. Technically, this means that the model includes main effects for school practice and context along with interaction terms for each of these two variables with student characteristics:

$$P_{ij} = Pj + (PS)_{ij} \text{ and } C_{ij} = Cj + (CS)_{ij}$$

With these equations as background, Raudenbush and Willms define Type A school effects in the following way:

$$A_{ij} = P_{ij} + C_{ij}$$

Here the effect of a school is made up of school practice (*P*) and the context in which the school resides (*C*). Type B school effects are defined in the following way:

$$B_{ij} = P_{ij}$$

Here, only the effects of school practice are considered. The differences between Type A and Type B effects are not trivial since one (Type A) includes the influence of environmental factors on student achievement, while the other does not. Although for many studies reviewed in this chapter it is difficult to ascertain specifically which type of school effect (i.e., A or B) has been addressed, in general it is safer to assume that discussions in the remainder of this chapter address Type A effects.

How Large Is the School Effect?

In Chapter 1 it was noted that the Coleman et al. (1966) study established the fact that schools account for about 10 percent of the variance of within-school achievement. Since then, a number of studies have attempted to identify the unique contribution of schools to student achievement. The results of some of the most prominent of these studies are reported in Table 4.2. In this section, not every study reported in Table 4.2 will be commented on — only those that have characteristics that provide a unique perspective on the effects of schools on student achievement.

The Coleman and Jencks reports are, of course, the studies of the effects of schooling that initially sparked an interest in (or, perhaps, the controversy over) the net impact of schooling. As mentioned in Chapter 1, the Jencks report used data collected for the Coleman report. Inspection of Table 4.2 indicates that these studies generated the lowest estimates of the effect size for schools. This discrepancy has been discussed in depth by Madaus, Kellaghan, Rakow, and King (1979). They note that although Coleman et al. had access to student scores on standardized tests of achievement in general information, reading, and mathematics, they used a general measure of verbal ability as the primary dependent measure. Additionally, this test primarily focused on vocabulary. This selection was made because Coleman and his colleagues found that the variation between schools was slightly greater for aptitude tests (i.e., verbal ability) than it was for achievement tests, thus providing "indirect evidence that variations among schools have as much or more effect on the ability scores as on achievement test scores" (p. 293). This use of general verbal aptitude as the primary dependent measure established a situation in which student background variables were highly likely to show much stronger relationships than were school-level variables. As explained by Madaus et al. (1979):

> Despite these difficulties with standardized tests, the construct "verbal ability" in the Coleman study has become equated with "school achievement" and the results have been generalized to the now popular myth that school facilities, resources, personnel, and curricula do not have a strong independent effect on achievement. Coleman's findings have been interpreted in the widest and most damaging possible sense, perhaps because verbal ability is considered so important, perhaps because of the tendency of social scientists to lose sight of the limits of their measures and to talk in broader and more commonly understood terms, and finally, perhaps because the media and public feel the need to simplify complex studies. To assert that schools bring little influence to bear on a child's general verbal ability that is independent of his background and general social context is not the same as asserting that schools bring little influence to bear on pupils' achievement in a specific college preparatory physics course. We might hope that schools would have some independent influence on general verbal ability. But the fact that home background variables seem to be vastly more influential in explaining verbal ability should not preclude or cloud any

> expectations we have that schools should have some independent effect on traditional curriculum areas which are systematically and explicitly treated as part of the instructional process. (p. 210)

In short, Coleman's choice of verbal ability as the primary dependent measure probably resulted in an underestimate of the effects of schooling on student achievement.

The effect size estimate by Byrk and Raudenbush (1992) is noteworthy in that it utilized a comparison between Type A and Type B effects. Using HLM on mathematics achievement data from 7,185 students nested in 160 schools, Byrk and Raudenbush estimated that school-level variables account for 18 percent of the variance (r = .42) in student achievement when the following model is used:

$$Y_{ij} = Bo_j + r_{ij}$$

Here, Y_{ij} is the achievement score for student i in school j. Bo_j is the average score for school j, and r_{ij} represent all those other factors that affect student achievement. In other words, the Byrk and Raudenbush model partials out all factors other than the school effect size into a large residual category (i.e., r_{ij}). However, when the average SES of schools was entered into the equation that has school mean as the outcome, Byrk and Raudenbush found that 69 percent of the variance is accounted for by SES. One might interpret this as an estimate of the Type B school effect since the average SES of students might be considered a good proxy measure of school context. If this is the case, then it indicates that Type B effects might be significantly lower than Type A. However, Teddlie, Reynolds, and Sammons (2000) provide evidence that certain HLM models can severely underestimate school-level effects. Specifically, they cite the HLM convention of "shrinking" residual values toward the mean as problematic from an interpretational perspective (p. 106).

Scheerens and Bosker (1997) provide still another perspective on the estimate of school effects. Using data from Bosker and Witziers (1995), they partitioned the school effects into two broad categories: gross effects and net effects. The gross school effects were based on the mean achievement scores for schools without corrections for any background variables such as SES of students, ethnicity, aptitude, and the like. Net school effects were based on the means of schools after the variance due to background variables had been accounted for. To determine the average gross and net school effects, Scheerens and Bosker examined findings from studies at the elementary and secondary levels that cut across three subject areas (language arts, mathematics, and science) in multiple countries (e.g., Netherlands, UK, other European countries, other industrialized countries, third-world countries). Using HLM, they examined the influences of studies and replications on gross and net school effects. (See Note 1 at the end of this chapter.) The percentage of variance accounted for by school membership for the gross school effect was 18.6. The percentage of variance accounted for by school membership for the net school effect was 8.4. However, when corrected for random "noise," the estimate of net school effect was raised to 11 percent.

Table 4.2
Summary of Studies on the Effect of Individual Schools on Student Achievement

Study	*ESd*	*P gain*	*PV*
Coleman et al. (1966)	.68 .80	25 29	10.38 13.89
Jencks et al. (1972)	.47 .56	18 21	5.29 7.29
Byrk & Raudenbush (1992)	.93	32	18.00
Scheerens & Bosker (1997)	.70	26	11.00
Rowe & Hill (1994)	1.32	40	30.00
Creemers (1994)	1.01	34	20.00
Stringfield & Teddlie (1989)	1.16	37	25.00
Bosker (1992)	1.19	38	26.00
Luyten (1994)	.85	30	15.00
Madaus et al. (1979)	1.04	35	21.84
$\bar{x}$ ($Q = 24.53$, $df = 9$, $p < .05$)	.96	33	18.49
$\bar{x}$ with outliers removed ($Q = 12.2$, $df = 7$, $p > .05$)	1.01	34	20

Note: Quantities were computed using data found in each of the studies listed in this table. Quantities were computed beginning with the *r* reported in each study. These were transformed to *Zr* and an average was computed. The average *Zr* was then transformed back to *r*. (See Note 4 at the end of this chapter for an explanation of how *Zr* was computed.) The *PV*, *ESd*, and *P gain* were then computed from this average *r*. The two effect sizes from the Coleman and Jencks studies were each given a weight of .5 when computing the average *r*. All other *r*'s were given a weight of 1.
r is the Pearson product-moment correlation; *PV* is percentage of variance explained; *ESd* is Cohen's *d*; *P gain* is percentile gain of experimental group. A *Q* statistic with $p < .05$ was interpreted as an indication that one or more correlations in the set were outliers. These outliers were identified using procedures described by Hedges and Olkin (1985). The *Q* statistic with outliers removed was then computed.

The effect size reported by Rowe and Hill (1994) is certainly much higher than most others reported in Table 4.2. This is probably because the dependent measures used in the Rowe and Hill study were experimenter-designed, open-ended tasks. The significance of the use of experimenter-designed dependent measures is discussed in more depth in the next section. Briefly, though, as discussed below, a strong case can be made that studies using experimenter-designed assessments might provide more valid estimates of school-level effects than do studies employing standardized assessments.

The school effect size estimate by Madaus et al. (1979) is unique because of its comparison of school effect size estimates based on standardized tests versus school effect estimates based on curriculum-specific assessments. Using data from Irish high schools, researchers were able to estimate the unique and common variance of a number of school-level variables and student-level variables. (See Note 2 at the end of the chapter for a discussion of the manner in which the effect size for this study was computed.) This was done using two sets of dependent measures — one set used standardized tests, the other used curriculum-specific tests designed to measure the content specific to the curriculum. The effect size reported in Table 4.2 is that computed using the curriculum-specific assessments. The Madaus et al. school effect size computed using standardized tests was *ESd* = .595, *PV* = 8.07, which is considerably smaller than that using curriculum-specific assessments. This discrepancy led Madaus et al. to note:

> Our findings provide strong evidence for the differential effectiveness of schools: differences in school characteristics do contribute to differences in achievement. The extent to which these differences can be detected is determined by the measure used. Examinations geared to the curricula of schools are more sensitive indicators of school performance than are conventional norm-referenced standardized tests. (p. 223)

The effect sizes reported in Table 4.2 lead to a different perspective on the effects of schools from that reported in the Coleman and Jencks reports. Specifically, the average effect size computed from Table 4.2 can be regarded as a viable estimate of the population effect size for schools. That average *ESd* is.96 with an associated *P gain* of 33 and *PV* of 18.49. However, Hedges and Olkin (1985) note that one might first ascertain the homogeneity (or lack thereof) of the set from which the average effect size is computed. If there are outliers in the set, the average will be biased in the direction of the outliers. Hedges and Olkin offer the Q statistic as an indicator of the homogeneity of effect sizes from which a given average effect size is computed. The Q statistic is distributed as chi square with (k-1) degrees of freedom where k is the number of effect sizes in the set. A significant (e.g., $p < .05$) statistic indicates that one or more elements of the set are outliers. Possible outliers can then be identified and removed until the Q statistic falls below the level of significance. As shown in Table 4.2, the Q statistic computed for the average *ESd* of .96 is significant ($p < .05$). When outliers are removed, the newly computed average *ESd* is 1.01 with an associated *P gain* of 34 and *PV* of 20.00. Again, the binomial effect size display (*BESD*) provides a useful way of interpreting this finding. The *BESD* of the new school effect size estimate is shown in Table 4.3.

Table 4.3 provides a practical interpretation of the new effect size estimate. Specifically, when the *PV* of schools is assumed to be 20.00, it implies that the percentage of students who would pass a state-level test (for which the expected passing rate is 50 percent) is 72.36 percent for effective schools versus 27.64 percent for ineffective schools, for a differential of 44.72 percent. This is not a trivial difference, especially for the 44.72 percent of students.

Table 4.3
Binomial Effect Size Display With School Accounting for 20% of Variance (r = .447)

Group	Outcome %		
	%Success	%Failure	Total
Effective Schools	72.36%	27.64%	100%
Ineffective Schools	27.64%	72.36%	100%

Note: r stands for the Pearson product-moment correlation coefficient.

The Case for Even Larger School-Level Effects

In this section an argument is presented that the effect of *some* schools might be even larger than that reported in Table 4.3. The assertion here is that the updated *PV* of 20 percent and its related effect size (*ESd*) of 1.01 might be an underestimate of the school effect, at least in some situations. Three lines of evidence support this assertion.

First, as Klitgaard and Hall (1974) argue, studies such as those reported in Table 4.2 focus on the *average* effect size of all schools in a given sample. Focusing on the average effect size ignores the fact that some schools will have effect sizes much larger than the average (and some schools will have effect sizes much smaller). As Klitgaard and Hall explain, even if one identifies the average effect size in the population, there still will be some highly effective schools whose effect sizes are much larger than the average.

To illustrate this point, it is useful to translate the average *ESd* of 1.01 reported in Table 4.2 into its equivalent correlation. Using the formula reported in Table 1.3, we compute the equivalent r to be .45. In other words, we can say that the average correlation of the studies reported in Table 4.2 is .45. Again, this is an average within a distribution of correlations. Knowledge of the variance of that distribution would provide us with information with which to estimate the extremes of the distribution.

One of the best estimates of the variance in the population of correlations from which the studies in Table 4.2 were chosen is that computed by Scheerens and Bosker. That variance is .0114. (See Note 3 at the end of this chapter.) If we assume that the correlations in the population of schools are distributed normally, then we can expect some schools to have correlations that are three standard deviations (or more) above the mean. In this case, the estimated standard deviation of the population of correlations is .1068 (i.e., $\sqrt{.0114}$). Consequently, one would expect some schools to have correlations three standard deviations above the mean, or .77 (.45 + .32). Reasoning from this perspective, one might make a case that the most effective of schools in the population could account for as much as 59.29 percent of the variance in student achievement ($.77^2 \text{x} 100 = 59.29$).

A second line of evidence to consider when examining the effect sizes in Table 4.2 is the fact that the dependent measures employed most commonly in these studies were some form of external standardized test. As mentioned previously, Madaus (Madaus et al., 1979; Madaus et al., 1980) has detailed the problems with this practice in terms of measuring the effectiveness of schools. Madaus et al. (1980) note that "one cannot . . . assume congruence between a commercially developed standardized test's objectives and those of a teacher" (p. 165). More pointedly, as Madaus et al. (1979) explain, the use of standardized tests as the primary dependent measure used to compute the school-level effect sizes creates some doubt about the validity of those estimates:

> Several of our results clearly indicate that what we call curriculum-sensitive measures are precisely that. Compared to conventional standardized tests, they are clearly more dependent on the characteristics of schools and what goes on in them. To have demonstrated this in one school system — any school system — is sufficient to cast serious doubt on the inferences drawn from other studies with their almost exclusive reliance on standardized, curriculum-insensitive tests — that schools do not differentially affect the attainments of their students. (pp. 223–224)

Commenting specifically on Coleman's findings, Madaus et al. (1979) note, "Had Coleman [and others] used measures which were more sensitive to the curriculum, would school factors have appeared more influential in explaining between-school variance? We feel the answer would be yes" (p. 225).

The final factor that supports the hypothesis that the effect size for some schools might be larger than r = .45 is the convention in the school effectiveness research to rarely, if ever, correct for the unreliability of the criterion measure — the assessment used as the indication of student achievement. Cohen and Cohen (1975) explain that random measurement error — unreliability of the measure — diminishes the size of the correlation between independent and dependent variables. They explain that it is reasonable to assume that as much as half of the variance in the criterion measures used in education research might be a function of random error due to the unreliability of these measures. To correct for attenuation due to unreliability, Hunter and Schmidt (1990) recommend that the following formula be used:

$$\boxed{\text{corrected } r = \frac{r}{\sqrt{\text{Reliability}}}}$$

Additionally, Jöreskog and Sörbom (1993) assert that .85 is the reliability one can reasonably assume for achievement and aptitude assessments. If one applies this correction to the average effect size (r) of .45 from Table 4.2, a corrected effect size of .48 is obtained.

In summary, the estimate of the school effect size used in the remainder of this monograph will be ESd = 1.01 with an associated r of .45, an associated PV of 20.00, and P *gain* of 34. However, a case can be made that there might be some "highly effective" schools with much larger effect sizes than the population average.

WHAT FACTORS ARE ASSOCIATED WITH THE SCHOOL EFFECT?

As described in Chapter 2, the model of school-level factors that emerged from the school effectiveness literature was a five-factor model (see Cohen, 1981; Odden, 1982; Ralph & Fennessey,

1983) that included the following:

1. Strong administrative leadership
2. A safe and orderly climate
3. An emphasis on basic academic skills
4. High expectations for student achievement
5. A system for monitoring pupil performance

Although the five correlates have intuitive appeal, their validity has been challenged. Commenting on these five factors, Willms (1992) notes:

> However, much of the literature on school process has been based on small comparative studies or ethnographies of exceptional schools. Critics claimed that the methods employed in these studies did not meet the standards of social science research; most studies did not control adequately for the background characteristics of students. . . . Although the five-factor model has considerable face validity, the empirical evidence that these factors are more important then some other set of factors is not compelling. (p. 327)

What, then, are the school-level variables that research indicates are most strongly related to student achievement and to what extent do they correspond to the "correlates?" Although the answers to these questions are still somewhat elusive, there is more of a research base with which these questions might be answered than there was in the 1970s. As mentioned in Chapter 3, the most quantitatively rigorous study to date of school variables was the meta-analysis by Scheerens and Bosker (1997), which built on previous studies by Bosker and Witziers (Bosker & Witziers, 1996; Witziers & Bosker, 1997). Given its breadth and rigor, it will be used as the basis for considering school-level variables.

Scheerens and Bosker utilized HLM to analyze the effect sizes. This allowed for the estimation of variance within studies and across studies (see Note 1 at the end of this chapter). The general findings reported by Scheerens and Bosker for school-level variables are summarized in Table 4.4.

In this section, we consider eight of these factors in more depth as possible candidates for the critical variables that constitute the school-level effect. More specifically, homework is excluded from the discussion here. It will be considered in Chapter 5 because research indicates that it is more of a teacher-level variable than a school-level variable (see Cooper, 1989).

Table 4.4
Effect Sizes from Scheerens and Bosker's Meta-Analysis

School-Level Variable	*N*	*Nr*	*Average ESd*[a]	*P gain*	*PV*
1. Cooperation	20	41	.0584	2	.08
2. School Climate	22	62	.2193	9	1.18
3. Monitoring	24	38	.2995	12	2.19
4. Content Coverage	19	19	.1767	7	.77
5. Homework	13	41	.1150	4	.29
6. Time	21	56	.3936	15	3.73
7. Parental Involvement	14	29	.2559	10	1.61
8. Pressure to Achieve	26	74	.2678	11	1.76
9. School Leadership	38	108	.0999	4	.25

Note: Data computed *from the Foundations of Educational Effectiveness*, page 305, by J. Scheerens and R. J. Bosker, 1997, New York: Elsevier.
N = number of studies. *Nr* = total number of replications across all studies, *ESd* is Cohen's *d*, *P gain* is the percentile gain of the experimental group, *PV* is the percentage of variance explained.
[a] Scheerens and Bosker report effect sizes using the Fisher Z transformation of zero-order correlations. (See Note 4 at the end of this chapter for an explanation of how *Zr* is computed.) The *Zr* was then transformed to *ESd*.

Cooperation

Cooperation has been identified by a variety of researchers as a school-level variable that impacts student achievement (see Venesky & Winfield, 1979; Glenn, 1991; Brookover & Lezotte, 1979; Frazer et al., 1987; Wang, Haertel & Walberg, 1993; and Cotton, 1995). At a very general level, cooperation can be described as the extent to which staff members in a school support one another by sharing resources, sharing ideas, and sharing solutions to common problems. Some indicators that signal cooperation at the school level are

- the frequency and quality of formal and informal meetings
- frequency and quality of informal contacts between staff
- the extent to which members agree on school policies
- the extent to which staff cooperation is an explicit goal
- the extent to which consensus is sought for critical decisions

As reported in Table 4.4, the average *ESd* for cooperation is .0584 with an associated *P gain* of 2 and *PV* of .08.

School Climate

School climate is a variable quite commonly cited in the research literature on school-level variables and one of the original five correlates (see Good & Brophy, 1986). It is defined here as the extent to which a school creates an atmosphere that students perceive as orderly and supportive. Indicators commonly associated with a positive school climate are

- clearly articulated and enforced rules and procedures
- orderly atmosphere
- positive interactions among staff and students
- implicit norms of civility are recognized and enforced

As Table 4.4 shows, the average *ESd* is .2193 with an associated *P gain* of 9 and *PV* of 1.18.

Monitoring

Monitoring refers both to the articulation of academic goals at the school level and the monitoring of progress toward those goals. Implicit in this variable is the collection of data on students' academic achievement and the use of those data to determine whether academic goals have been met. To monitor progress relative to academic goals, one must have access to student achievement data.

Again, this school-level variable can be considered one of the original correlates or strongly related to one of the original correlates (Good & Brophy, 1986). Some specific behaviors that indicate effective monitoring include the following:

- A strong emphasis on using assessment results to determine how well students are learning critical content.
- Basing instructional decisions on judgments about student learning.
- Comparing results of student assessment based on standardized or state-level assessments with those at the classroom level.

The average *ESd* for monitoring is .2995 with an associated *P gain* of 12 and *PV* of 2.19.

Content Coverage

As reported in Table 4.4, the average *ESd* for this variable is .1767 with an associated *P gain* of 7 and *PV* of .77. As defined in the Scheerens and Bosker (1997) analysis, content coverage includes factors such as

- ensuring that the curriculum is well articulated, and
- monitoring the extent to which the curriculum is addressed by classroom teachers.

It should be noted that this description does not include the extent to which the content addressed in the curriculum covers the content on which students are assessed. In the days of the school effectiveness research, the term "curriculum/test congruence" was sometimes used to reflect this variable. Specifically, curriculum/test congruence addresses the issue of coverage of content on the test. Without relatively high curriculum/test congruence, a school whose curriculum is well covered might, in fact, help students learn, but those students might not be learning the content covered by the test used as the criterion measure for student achievement.

The concept that the curriculum students are taught should mirror the assessments by which student achievement is judged and vice versa is strongly associated with the concept of "opportunity to learn" or OTL (Kifer, 2000). Creemers has reviewed many of the studies on the relationships between OTL and student achievement. These findings are summarized in Table 4.5.

Table 4.5
Results for Opportunity to Learn (OTL)

Study	*ESd*	*P gain*	*PV*
Husen, 1967	.68	25	10.24
Horn & Walberg, 1984	1.63	45	39.69
Pelgrum et al., 1983	.45	17	4.84
Bruggencate et al., 1986	1.07	36	22.09
$\bar{x}$ ($Q = 33.02$, $df = 3$, $p < .05$)	.94	33	18.06
$\bar{x}$ ($Q = 3.45$, $df = 1$, $p > .05$)	.88	31	16.00

Note: Statistics reported in this table computed from data presented in *The Effective Classroom, by B. P. M.* Creemers, 1994, London: Cassell. Quantities were computed by beginning with the *r* reported in each study. These were transformed to *Zr* and an average was computed. The average *Zr* was then transformed back to *r*. The *PV*, *ESd*, and *P gain* were then computed from the average *r*.
r is Pearson's product-moment correlation, *PV* is percentage of variance explained, *ESd* is Cohen's *d*, and *P gain* is percentile gain of experimental group.
A *Q* statistic with $p < .05$ was interpreted as an indication that one or more correlations in the set were outliers. These outliers were identified using procedures described by Hedges and Olkin. The *Q* statistic with outliers removed was then computed.

The effect sizes reported in Table 4.5 are quite high compared to those for curriculum coverage reported in Table 4.4. In fact, the average *r* with outliers removed is .400 with an associated *PV* of 16.00, *ESd* of .88, and *P gain* of 31. The strength of the OTL relationship with student achievement and its logical appeal make it a more useful school-level variable in terms of explaining the effects of schooling on student achievement than content coverage. Consequently, for the remainder of this monograph, the variable OTL will replace Scheerens and Bosker's variable content coverage. This variable will be defined as the extent to which a school (1) has a well-articulated curriculum, (2) addresses the content in those assessments used to make judgments about student achievement, and (3) monitors the extent to which teachers actually cover the articulated curriculum.

Time

One of the most enduring school-level factors in the research literature is the effective use of time (see Berliner, 1979). As Table 4.4. shows, the average *ESd* is .3936 with an associated *P gain* of 15 and *PV* of 3.73. The effect of time on achievement is, by far, the strongest identified in the Scheerens and Bosker (1997) study.

In the context of the Beginning Teacher Evaluation Studies (see Denham & Lieberman, 1980), the effects of time were studied in great depth. Specifically, time was classified in those studies into four basic types: allocated time, instructional time, engaged time, and academic learning time (Borg, 1980). Allocated time is that time in the school day specifically set aside for instruction, such as classes, as opposed to noninstructional activities, such as recess, lunch, passing time, and the like. Instructional time is the in-class time that a teacher devotes to instruction as opposed to management-oriented activities. Engaged time is that portion of instructional time during which students are actually paying attention to the content being presented. Finally, academic learning time is the proportion of engaged time during which students are successful at the tasks they are engaged in. Each of these categories of time has a stronger relationship with achievement than the previous type. In other words, academic learning time has a stronger relationship with achievement than does engaged time, and so on.

Although Scheerens and Bosker do not explicitly describe the type of time they are referring to, one can infer from their comments that they are not considering engaged time or academic learning time. Rather, it appears that the variable time does not go beyond allocated time. Stated differently, the variable of time as defined by Scheerens and Bosker includes

- maximizing the amount of time allocated for instruction,
- minimizing the amount of instructional time lost to absenteeism and tardiness, and
- minimizing the amount of instructional time lost to unnecessary extracurricular activities.

Parental Involvement

Parental involvement can be described in general terms as the extent to which parents are involved in and supportive of the culture and operating procedures of the school. It is a variable that was not highlighted as important within the school effectiveness movement. To illustrate, commenting on parental involvement within the school effectiveness literature, Good and Brophy (1986) note:

> The degree of home and school cooperation is likely to be an important determinant of student achievement. However, this “obvious” possibility has received little research attention. Whether parent-school communication differs in “more” and “less” effective schools is also unclear. (p. 590)

As indicated in Table 4.4, the average *ESd* for this variable is .2559 with an associated *P gain* of 10 and *PV* of 1.61. Some of the specific behaviors that constitute this factor are

- good written information exchange between school and parents,
- parental involvement in policy and curricular decisions, and
- easy access for parents to administrators and teachers.

Pressure to Achieve

Pressure to achieve in the Scheerens and Bosker study is basically synonymous with the school effectiveness correlate of high expectations for student achievement. It can be defined as the communication of a strong school-level message that academic achievement is one of the primary goals of the school. Specific behaviors within this category include the following:

- A clear focus on mastery of basic subjects.
- High expectation for all students.
- Use of records of student progress.

The average *ESd* for this category is .2678 with an associated *P gain* of 11 and *PV* of 1.76.

School Leadership

School leadership is defined here as the extent to which the school has strong administrative leadership relative to the goal of academic achievement. The factors associated with effective leadership defined in this way are

- well-articulated leadership roles,
- the school leader is an information provider, and
- the school leader facilitates group decision making.

The average *ESd* for this factor is .0999 with an associated *P gain* of 4 and *PV* of .25. This is the smallest effect size of the eight factors identified by Scheerens and Bosker, which is somewhat surprising since strong administrative leadership is one of the five correlates in the effective schools literature. One reason for the relatively small effect size computed by Scheerens and Bosker might be the way that school leadership is defined in their study as opposed to how it is defined in the school effectiveness literature. In the Scheerens and Bosker study, leadership focuses primarily on "quality control." In the school effectiveness literature, the definition of strong administrative leadership goes well beyond this function. In fact, one might argue that in the school effectiveness literature, leadership from the principal encompasses a majority of the school-level variables identified by Scheerens and Bosker (see Good & Brophy, 1986; Manassee, 1985). Specifically, school leadership as defined in the school effectiveness literature encompasses functions such as establishing policies relative to the use of time, establishing policies relative to curriculum/test congruence, and the like.

Conclusions about the School-Level Variables

If one accepts the interpretations just discussed, a rather straightforward picture emerges about the school-level variables that affect student achievement and their relative influences. Specifically, the eight factors drawn from the Scheerens and Bosker study might be ordered from largest effect to smallest as shown in Table 4.6. In addition, these eight factors might be compared to the five correlates from the school effectiveness literature as shown in Table 4.7.

Table 4.6
School-Level Variables

Variable	*ESd*	*P gain*	*PV*
Opportunity to Learn	.88	31	16.00
Time	.39	15	3.61
Monitoring	.30	12	2.19
Pressure to Achieve	.27	11	1.76
Parental Involvement	.26	10	1.61
School Climate	.22	8	1.18
Leadership	.10	4	.25
Cooperation	.06	2	.08

Note: PV is percentage of variance explained, *ESd* is Cohen's *d*, and *P gain* is percentile gain of experimental group.

Table 4.7
Comparison with School Effectiveness Correlates

School Effectiveness Correlates	Scheerens and Bosker Variables
• Administrative leadership	• Cooperation • School leadership
• Safe and orderly climate	• School climate
• Emphasis on basic skills	• Opportunity to learn
• High expectations	• Pressure to achieve
• Monitoring pupil performance	• Monitoring
	• Parental involvement • Time

As Table 4.7 shows, at least six of the eight variables considered in this chapter can be thought of as strongly related to or identical to the five school effectiveness correlates, with the only outliers being time and parental involvement.

Another point that should be made about the school-level factors that make up the school effect is their relatively small effect sizes, which should not be misconstrued as an indication that these variables are not important. As illustrated in Chapter 1 using the *BESD* (binomial effect size display), relatively small effect sizes can have a rather profound effect on student achievement. Further, as described by Brophy and Good (1986), "Many of the school effects variables probably have nonlinear relationships with outcomes" (p. 588). This would imply that factors like parental involvement, for example, have a positive influence on student achievement up to a certain point, after which an increase in this variable no longer affects student achievement or might influence it negatively.

The final conclusion that should be noted from this chapter is the updated estimate of the school-level effect. Specifically, that estimate is an *r* of .45 with an accompanying *PV* of 20.00, *ESd* of 1.01, and *P gain* of 34.

Chapter 4 Notes:

Note 1:
The within-replication model used by Scheerens and Bosker (1997) was $d_{rs} = \delta_{rs} + e_{rs}$, indicating that the effect size in replication *r* of study s is made up of an estimate of the population effect size for a replication within a given study (δ_{rs}) plus an error component due to sampling error (e_{rs}). The between-replication model was $\delta_{rs} = \delta_s + u_{rs}$, indicating that the average effect size for all replications within a given study is made up of an estimate of the population effect size for the replications within the study (δ_s) plus an error component due to sampling errors (u_{rs}).

Finally the between-studies model was $\delta_s = \delta_o + v_s$, indicating that the effect size estimate within a given study is comprised of the overall population effect size (δ_o) plus an error component. Thus, the effect computed for a specific replication (d_{rs}) can be represented in the following way:

$$d_{rs} = \overbrace{\underbrace{\delta_o + v_s}_{}}^{\delta_s} \!\!\!\!\!\!\!\!\!\!\!\!\!\! $$

$$d_{rs} = \underbrace{\overbrace{\delta_o + v_s}^{\delta_s} + u_{rs}}_{\delta_{rs}} + e_{rs}$$

It is also important to note that Scheerens and Bosker used an effect size estimate (Cohen's *F*) that was appropriate to address the variation in multiple means (i.e., ANOVA designs) as opposed to an effect size estimate based on a comparison of two means (e.g., Cohen's *d*). Cohen's *F* is defined as

$$\boxed{F = \sqrt{\frac{P}{1-P}}}$$

where P is the interclass correlation coefficient operationally defined as the percentage of the total variance explained by the variance between groups — in this case, the variance between schools. Although Cohen's *F* ranges from 0 to infinity, up to the value of about .50 it is roughly equivalent to *r* in terms of its interpretation.

The results of Scheerens and Bosker's analysis are reported as follows:

Gross and Net School Effects

	Gross School Effect	Net School Effect
Average Effect	.4780	.3034
Variance Across Studies	.0332	.0111
Variance Across Replications	.0070	.0003

Note: Data from *The Foundations of Educational Effectiveness*, (pp. 76 and 78), by J. Scheerens and R. J. Bosker, 1997, New York: Elsevier.

As this table shows, the variance accounted for across replications is negligible, but the variation across studies is not for both gross and net effects. The combined variance across studies and within studies of .0114 (.0111 + .003) provides a useful estimate of the variance one might expect in various estimates of the net school effect when these estimates are expressed as zero-order correlations.

Note 2:
The effect size for the Madaus et al. study (1979) was computed by taking the average of the unique variances for the "classroom" plus the "individual classroom" variables as reported in Table 3, page 219 of that study for curriculum-specific measure. This average variance was then transformed to *r*.

Note 3:
To compute this standard deviation, the variance across studies and across replications reported in the table shown in Note 1 was summed.

Note 4:
Fisher's *Zr* is computed using the following formula:

$$Zr = \frac{1}{2} \log_e \frac{1 + r}{1 - r}$$

In general, when *r* is small (e.g., < .2) *r* and *Zr* are very close in value. But when *r* is larger than .2, *Zr* will be larger than its corresponding *r*.

Chapter 5
THE TEACHER-LEVEL EFFECT

In this chapter we consider those variables that are specific to individual teachers within a school as well as the overall effect of the teacher. Stated differently, we consider those variables that are under the control of individual teachers regardless of the context provided by the school — those things a teacher might do to enhance student achievement no matter what the school's position is about monitoring student achievement, providing a positive climate, and so on. Brophy and Good (1986) describe the need to address teacher-level effects separately from school-level effects as follows:

> Studies of large samples of schools yield important profiles of more and less successful schools, but these are usually *group averages* that may or may not describe how a single effective teacher actually behaves in a particular effective school. Persons who use research to guide practice sometimes expect all teachers' behavior to reflect the group average. Such simplistic thinking is apt to lead the literature to be too broadly and inappropriately applied. (p. 588)

In short, this chapter seeks to answer the questions, How big is the teacher-level effect? and What constitutes that effect?

HOW BIG IS THE TEACHER-LEVEL EFFECT?

Most of the research on school-level effects discussed in Chapter 4 "sums over" the effect of teachers within a specific school. The effect of an individual teacher, then, is lost in the average for the school. In this chapter, we first try to separate the effects of an individual teacher from that of a school. Scheerens and Bosker (1977) note that unless the teacher effect is separated from the school effect

> the [school] effect size is overestimated, since the important intermediate level of the classroom is ignored. . . . In general, ignoring the intermediate classroom level leads to an overestimate of school effects. This overestimate amounts to variance between classes within schools divided by the average number of classes within schools. (p. 80)

Results from the more salient studies that have attempted to partial out the teacher effect from the school effect are reported in Table 5.1. The quantities reported in this table reveal a somewhat inconsistent picture of the relative effects of schools versus teachers. In the Springfield and Teddlie (1998) and Bosker (1992) studies, the percentage of variance accounted for by school variables and teacher variables is about equal. However, the Luyten (1994) study ascribes twice as much of an effect to teachers as it does to schools. The Madaus et al. (1979) study addresses the issue of unique contributions of schools versus teachers perhaps most directly. As mentioned in Chapter 4, in this study the unique and common variance for schools versus teachers was computed on a number of dependent measures. The figures reported in Table 5.1 for the Madaus et al. study indicate that the ratio of teacher to school effect is about 4.5 to 1 (i.e., 18 to 4) — the largest ratio of teacher to school effects among the studies reported in this table.

Table 5.1
The Teacher-Level Effect

Study	Percentage of Variance		
	School & Teacher	School	Teacher
Stringfield & Teddlie (1989)[a]	25%	13%	12%
Bosker (1992)[a]	25% (math) 27% (language)	15% 13%	11% 14%
Luyten (1994)[a]	15%	5%	10%
Madaus et al. (1979)[b]	22%	4%	18%

[a] See these studies for data reported in this table.
[b] Data computed from data found in Madaus et al. (1979). These researchers report the unique and common variance for two school-level factors that they refer to as the "classroom block" and the "individual/classroom" block. Despite the use of the term "classroom" to describe both categories of variables, the classroom block category is closely related to the teacher-level variables as described in this monograph, and the individual/classroom category is most closely related to the school-level variables as described in this monograph. The average unique variance for scores on curriculum-specific dependent measures was used as the estimate of the variance accounted for by these two categories of variables.

There is some rather compelling research evidence that cannot easily be interpreted in effect size metrics. This evidence supports the assertion that the effects of teachers far exceed the independent effects of schools. Specifically, the primacy of the teacher effect over the school effect has been firmly established by Sanders and his colleagues within the context of the Tennessee Value-Added Assessment System (TVAAS) (see Sanders & Horn, 1994; Wright, Horn, & Sanders, 1997).

Reporting on 30 separate analyses across three grade levels (3–5) and five subject areas (math, reading, language arts, social studies, science) with some 60,000 students, Wright et al. (1997) found a number of interesting patterns. Specifically, the TVAAS researchers utilized the convention of computing p-values for each F statistic and then translating each p value to its corresponding z-score by treating the p-values as two-tailed, standard normal deviates. Consequently, .05, .01, .001, and .0001 levels of significance correspond to Z scores of 1.96, 2.58, 3.29, and 3.89, respectively. The general findings from Wright et al.'s analysis are reported in Table 5.2.

Perhaps most striking about Table 5.2 is the consistently significant effects of teachers. The effect of the teacher was significant at the .0001 level 100 percent of the time. This is particularly compelling inasmuch as 30 separate estimations were computed for each factor. No other factor had this level of consistency in the findings, not even the prior achievement of students (A). Another interesting finding reported in Table 5.2 is that the heterogeneity of the class was significant in only 3.3 percent of the 30 contrasts at the .05 level and not at all at higher levels of significance.

Table 5.2
Findings From Tennessee Value Added System (TVAAS) Studies

Factor	Level of Significance			
	≤ .05 (1.96)[a]	≤ .01 (2.58)[a]	≤ .001 (3.29)[a]	≤ .0001 (3.89)[a]
School (S)	27/30 = 90%	24/30 = 80%	20/30 = 66.7%	16/30 = 53.3%
Heterogeneity (H)	1/30 = 3.3%	0/30 = 0%	0/30 = 0%	0/30 = 0%
Class Size (C)	3/30 = 10%	1/30 = 3.3%	0/30 = 0%	0/30 = 0%
H*C	4/30 = 13.3%	0/30 = 0%	0/30 = 0%	0/30 = 0%
Teacher (S*H*C) (T)	30/30 = 100%	30/30 = 100%	30/30 = 100%	30/30 = 100%
Achievement Level (A)	26/30 = 86.7%	23/30 = 76.7%	23/30 = 76.7%	21/30 = 70%
A*S	21/30 = 70%	14/30 = 46.7%	8/30 = 26.7%	3/30 = 10%
A*H	10/30 = 33.3%	5/30 = 16.7%	4/30 = 13.3%	2/30 = 6.7%
A*H*C	4/30 = 13.3%	1/30 = 3.3%	0/30 = 0%	0/30 = 0%
A*T	9/30 = 30%	4/30 = 13.3%	2/30 = 6.6%	1/30 = 3.3%

Note: Table constructed using data from "Teacher and Classroom Context Effects on Student Achievement. Implications for Teacher Evaluation" (pp. 60–62), by S. P. Wright, S. P. Horn, and W. L. Sanders, 1997, *Journal of Personnel Evaluation in Education, 11,* 57–67.
Heterogeneity (*H*) refers to the variance in achievement of students within a given class.
The term *T* refers to the effects of individual teachers nested within a particular school (*S*), within a class with a specific level of heterogeneity (*H*), with a specific class size (*C*).
The term *A* stands for the average prior achievement of students within a class. All other terms in this table are interpreted in the traditional manner for interactions.
[a] *Z* score

These results lead Wright et al. (1997) to note:

> The results of this study will document that the most important factor affecting student learning is the teacher. In addition, the results show wide variation in effectiveness among teachers. The immediate and clear implication of this finding is that seemingly more can be done to improve education by improving the effectiveness of teachers than by any other single factor. *Effective teachers appear to be effective with students of all achievement levels regardless of the levels of heterogeneity in their classes*. If the teacher is ineffective, students under that teacher's tutelage will achieve inadequate progress academically, regardless of how similar or different they are regarding their academic achievement. (p. 63) [emphases in original]

In sum, it appears safe to conclude that the variance accounted for by the individual classroom teacher is greater than that accounted for uniquely by the school as a unit. Exactly what unique percentage of variance to ascribe to schools versus teachers is not clear. However, a realistic yet somewhat conservative estimate appears to be a ratio of 2 to 1 in favor of teachers. In the remainder of this monograph, we will assume that of the 20 percent of variance accounted for by schools as concluded in Chapter 4, 13.34 percent is a function of teacher-level variables and 6.66 percent is a function of school-level variables.

What Constitutes the Teacher-Level Effect?

As is the case with school-level variables, lists of teacher-level variables abound in the research literature. For example, Cotton (1995) lists more than 160 teacher-level variables that contribute to student achievement. Frazer et al. (1987) list 25 variables; Walberg (1999) lists some 30 variables; and Scheerens (1992) lists more than 30. In spite of the exhaustive lists of teacher-level variables, three categories are commonly used to organize them: (1) instruction, (2) curriculum design, and (3) classroom management.

Instruction

The category of instruction is defined here as including those direct and indirect activities orchestrated by the teacher to expose students to new knowledge, to reinforce knowledge, or to apply knowledge. Within this category, Creemers (1994) lists the following:

- Advance organizers
- Evaluation
- Feedback
- Corrective instruction
- Mastery learning
- Ability grouping
- Homework
- Clarity of presentation
- Questioning

In a meta-analysis of research on instruction, Marzano (Marzano, 1998; Marzano, Gaddy, & Dean, 2000; Marzano, Pickering, & Pollock, 2001) identified nine categories of instructional variables. These are reported in Table 5.3 along with their effect sizes.

It is important to comment on the relatively large effect sizes reported in Table 5.3 for specific instructional strategies as opposed to those reported in Table 4.2 for the general effects at the school level. The reason for this disparity has already been addressed in another context. Specifically, the effect sizes in Table 5.3 are much higher than those reported in Table 4.2 because the studies from which the effect sizes in Table 5.3 were computed used assessments that were specifically designed to assess the dependent variable in question. That is, the assessments used in these studies were "experiment-specific." As described previously, Madaus et al. (1979) have shown that assessments specific to the curriculum being taught are far more sensitive to effects due to schools, teachers, or both than more general standardized tests.

Table 5.3
Nine Categories of Instructional Strategies

Category	*ESd*	*P gain*	*P V*
Identifying similarities and differences	1.61	45	27.04
Summarizing and note taking	1.00	34	20.25
Reinforcing effort and providing recognition	.80	29	13.69
Homework and practice	.77	28	12.96
Nonlinguistic representations	.75	27	12.25
Cooperative learning	.73	27	11.56
Setting goals and providing feedback	.61	23	8.41
Generating and testing hypotheses	.61	23	8.41
Activating prior knowledge	.59	22	7.84

Note: ESd is Cohen's *d*; *P gain* is percentile gain of experimental group; *PV* is percentage of variance explained.

Marzano (Marzano, 1998; Marzano et al., 2000; Marzano et al., 2001) also has organized the nine instructional categories reported in Table 5.3 into a sequence for "unit design" as shown in Table 5.4. This protocol combines the nine instructional categories into a planning framework for units as opposed to individual lessons as was the case with the planning framework Hunter (1984) designed.

Curriculum Design

The category referred to as curriculum design addresses the order and pacing of content and instructional activities. To distinguish this category of variables from those in the category of instruction, consider the fact that a teacher could use all of the instructional strategies listed in Table 5.3, but still not address the subject-matter content in a logical way or pace activities in a way that optimizes learning.

Creemers lists two factors in this category: (1) explicit ordering of goals, and (2) clearly stated and well-structured content. These factors are brought to life in the context of Bloom's (1976) research on the nature and structure of classroom tasks. Bloom reasoned that during a year of school, students encounter about 150 separate "learning units or learning tasks" (p. 87), each representing about seven hours of school work. Assuming that the school day is divided into five academic courses, we can infer that students encounter about 30 learning units within a year-long course or about 15 learning units within a semester-long course. What is referred to here as curriculum design might be operationally defined as the extent to which activities within these learning units are organized in a way that optimizes learning and the extent to which learning units are ordered in a way that optimizes learning. According to Clark and Yinger (1979), this aspect of instruction also involves selecting appropriate learning activities and organizing these activities within and between units.

Table 5.4
Planning Guide

When Strategies Might be Used	Instructional Strategies
At the Beginning of a Unit	***Setting Learning Goals*** 1. Identify clear learning goals. 2. Allow students to identify and record their own learning goals.
During a Unit	***Monitoring Learning Goals*** 1. Provide students feedback and help them self-assess their progress toward achieving their goals. 2. Ask students to keep track of their achievement of the learning goals and of the effort they are expending to achieve the goals. 3. Periodically celebrate legitimate progress toward learning goals. ***Introducing New Knowledge*** 1. Guide students in identifying and articulating what they already know about the topics. 2. Provide students with ways of thinking about the topic in advance. 3. Ask students to compare the new knowledge with what is known. 4. Have students keep notes on the knowledge addressed in the unit. 5. Help students represent the knowledge in nonlinguistic ways, periodically sharing these representations with others. 6. Ask students to work sometimes individually, but other times in cooperative groups. ***Practicing, Reviewing, and Applying Knowledge*** 1. Assign homework that requires students to practice, review, and apply what they have learned; however, be sure to give students explicit feedback as to the accuracy of all of their homework. 2. Engage students in long-term projects that involve generating and testing hypotheses. 3. Have students revise the linguistic and nonlinguistic representations of knowledge in their notebooks as they refine their understanding of the knowledge.
At the End of a Unit	***Helping Students Determine How Well They Have Achieved Their Goals*** 1. Provide students with clear assessments of their progress on each learning goal. 2. Have students assess themselves on each learning goal and compare these assessments with those of the teacher. 3. Have students articulate what they have learned about the content and about themselves as learners.

Research by Nuthall (Nuthall, 1997; Nuthall & Alton-Lee, 1995) provides some guidance for within-unit and between-unit planning. Specifically, Nuthall's research indicates that students should be exposed to informational knowledge at least three or four times before they can legitimately be expected to remember that information or use it in meaningful ways. In addition, the time between exposures to that information should not exceed about two days. The interval created by the need for multiple exposures to information and the need for those exposures to be relatively close in time has been called the "time window" for learning (Rovee-Collier, 1995).

Also relevant to this discussion is Kulik and Kulik's (1989) meta-analysis of the effects of goal structure on student achievement. Specifically, they report an effect size (*ESd*) of .30 when goals are well articulated and organized into a hierarchical structure. Finally, Creemers (1994) makes the following comment about the structure of goals and their influence on student achievement:

> The hierarchy of goals is reflected in the structure of a curriculum starting with easy exercises and simple knowledge and building up to more complex exercises and knowledge structures . . . Research shows that clearly structured curricula are more effective than less clearly structured curricula. The clear structure is expressed in goals that should be achieved in succession: achieving the first goal is a condition for achieving later goals. (p. 49)

In summary, effective curriculum design appears to be a function of the learning goals that are established by the teacher, the manner in which these goals are organized, the activities selected to help students meet these goals, and the manner in which these activities are spaced and paced.

Classroom Management

Classroom management involves those teaching behaviors and teacher designed activities that are designed to minimize disruptions or distractions to the learning process and maximize the effectiveness of interaction between teachers and students, and students and students. It is certainly noteworthy that in their analysis of 30 variables influencing student achievement, Wang et al. (1993) listed classroom management as the most influential. (See Chapter 3 for a discussion.) Again, the lists of factors within this instructional category can be quite long. Cotton (1995) lists 19 factors that deal with management; Scheerens and Bosker (1997) list 22 elements.

In much of the research literature, the classroom management variables overlap greatly with variables in the previous two categories — instruction and curriculum design. This makes intuitive sense — well-planned units that use the most effective instructional strategies will require little attention to management. However, some unique classroom management activities have been identified by Emmer et al. (1984) and Evertson et al. (1984). These are reported in Table 5.5 for elementary and secondary classrooms.

As Table 5.5 shows, classroom management involves establishing and implementing procedures and rules for routine and nonroutine activities in the day-to-day life of the classroom. Although there certainly are differences between management concerns in the elementary and secondary classroom, both have a great deal in common including establishing and implementing procedures and rules for seat work, group work, and discipline.

Table 5.5
Classroom Management Variables

Elementary School Management	*Secondary School Management*
Room use: • teacher's desk and storage • student's desk and storage • bathroom use • use of centers and stations	Seat work: • student attention • student participation • student talk • out-of-seat behavior • when seatwork is completed
Seat work: • student attention and participation • talking during seatwork • obtaining help • out-of-seat procedures • activities after seatwork is completed	Group work: • different roles in groups • use of materials • student participation and behavior
Group work: • group behavior • individual behavior within a group	Discipline: • loss of privileges • checks or demerits • detention • restitution • confiscation
Discipline: • loss of privileges • checks or demerits • detention • restitution • confiscation	General procedures: • distributing material • behavior during disruption • special equipment
General procedures: • distributing material • interrupting • fire and disaster drills • classroom helpers	

Note: See *Classroom Management for Secondary Teachers*, by E. T. Emmer, C. M. Evertson, J. P. Sanford, B. S. Clements, and M. E. Worsham, M. E, 1984, Englewood Cliffs, NJ: Prentice Hall; and *Classroom Management for Elementary Teachers*, by C. M. Evertson, E. T. Emmer, B. S. Clements, J. P. Sanford, and M. E. Worsham, 1984, Englewood Cliffs, NJ: Prentice Hall.

CONCLUSIONS ABOUT TEACHER-LEVEL VARIABLES

Based on the research on the effects of teacher-level variables, one can conclude that a reasonable estimate of the relative effects of teachers versus schools is 2 to 1. Chapter 4 established that a viable estimate of the effects of schooling is that it accounts for about 20 percent of the variance in student achievement. Thus, 13.34 percent can be assigned to teachers and 6.66 percent to schools using the 2 to 1 ratio. In addition, as described in this chapter, the unique effects of individual teachers can be thought of as consisting of the effective use of specific instructional strategies, effective curriculum design, and effective classroom management.

Chapter 6
THE STUDENT-LEVEL EFFECT

One of the perceived "truisms" in education is that students' background characteristics account for the lion's share of the variation in student achievement. Again, this was one of the primary conclusions of the Coleman et al. (1966) and Jencks et al. (1972) reports. In keeping with the two preceding chapters, this chapter addresses the questions, How big is the student effect? and What constitutes that effect?

HOW BIG IS THE STUDENT-LEVEL EFFECT?

An assumption not uncommon in the school effectiveness research is that all variances that cannot be accounted for by school- and classroom-level characteristics can be attributed to named or unnamed student-level variables. This convention is used in this monograph — the overall student-level effect is computed from the overall school effect. To illustrate, Table 6.1 contains the student-level effects as computed from Table 4.2.

Table 6.1
Estimates of Student-Level Effect

Study	*ESd*	*P gain*	*PV*
Coleman et al. (1966)	5.89 4.98	>49 >49	89.62 86.11
Jencks et al. (1972)	8.43 7.15	>49 >49	94.71 92.71
Byrk & Raudenbush (1992)	4.28	>49	82.00
Scheerens & Bosker (1997)	5.67	>49	89.00
Rowe & Hill (1994)	3.06	>49	70.00
Creemers (1994)	4.00	>49	80.00
Stringfield & Teddlie (1989)	3.49	>49	75.00
Bosker (1992)	3.38	>49	74.00
Luyten (1994)	4.76	>49	85.00
Madaus et al. (1979)	3.71	>49	78.16
$\bar{x}$ =	3.92	>49	80.00

Note: Averages were computed from Table 4.2. Specifically, the average *PV* with outliers excluded was subtracted from 100 to compute the *PV*. *r* was then computed from *PV*; *ESd,* and *P gain* were computed from *r*.
r is Pearson's product-moment correlation; *PV* is percentage of variance explained; *ESd* is Cohen's *d*; *P gain* is percentile gain of experimental group.

It should be noted that the approach taken in Table 6.1 is highly conservative in terms of the effects of schools and teachers. That is, it gives the benefit of the doubt to factors outside of the influence of the school or classroom. This approach is used in this monograph in order to avoid drawing overly optimistic conclusions about the potential of school reform. Stated differently, this monograph seeks to demonstrate that even the most conservative perspective on the effects of schools and classrooms on student achievement still indicates that schools and teachers can have a profound effect on student achievement.

WHAT CONSTITUTES THE STUDENT-LEVEL EFFECT?

As is the case with school and teacher levels, there is no single way to organize the research on student-level variables. However, four factors are commonly considered in discussions of student background — socioeconomic status (SES), prior knowledge, interest, and aptitude.

Socioeconomic Status (SES)

According to White (1982), the Coleman report confirmed for educators what they thought they already knew — "that a strong relationship exists between all kinds of academic achievement variables and what has come to be known as socioeconomic status (SES)" (p. 46). White notes that the belief in the strong relationship between SES and achievement is so prevalent in the research literature that it is rarely questioned. As proof, he offers the following set of quotes:

> The family characteristic that is the most powerful predictor of school performance is socioeconomic status (SES): the higher the SES of the student's family, the higher his academic achievement. This relationship has been documented in countless studies and seems to hold no matter what measure of status is used (occupation of principal breadwinner, family income, parents' education, or some combination of these). (Boocock, 1972, p. 32)
>
> To categorize youth according to the social class position of their parents is to order them on the extent of their participation and degree of success in the American Educational System. This has been so consistently confirmed by research that it can now be regarded as an empirical law. . . . SES predicts grades, achievement and intelligence test scores, retentions at grade level, course failures, truancy, suspensions from school, high school dropouts, plans for college attendance, and total amount of formal schooling. (Charters, 1963, pp. 739–740)
>
> The positive association between school completion, family socioeconomic status, and measured ability is well known. (Welch, 1974, p. 32)

White argues that in spite of the testimonies to the strong relationship between SES and academic achievement, reported correlations do not paint a clear picture. Specifically, correlations range from .10 to .80 as reported in the research literature. White speculates that one factor contributing to the variation in reported relationships between SES and achievement is the variation in the way SES is defined and, consequently, measured. In a meta-analysis of 101 reports yielding 636 effect sizes, White found the pattern of results reported in Table 6.2. (It should be noted that White reported his findings in terms of *r*; in Table 6.2, these statistics have been translated to *ESd*.)

Table 6.2
Effects of Various Aspects of SES on Achievement

SES Indicator	*ESd*	*P gain*	*PV*
Income only	.67	25	9.92
Education only	.38	24	3.24
Occupation only	.42	26	4.04
Home atmosphere only	1.42	42	33.29
Income and education	.47	18	5.29
Income and occupation	.70	26	11.02
Education and occupation	.69	26	10.56
Income, education, and occupation	.66	25	10.11

Note: Data used to calculate the numbers presented in this table are from "The Relationship Between Socioeconomic Status and Academic Achievement," by K. R. White, 1982, *Psychological Bulletin, 91*(3), p. 470. White's original findings were reported in terms of *r*. *ESd*, *P gain*, and *PV* were computed from the reported *r*'s. *r* is Pearson's product-moment correlation; *PV* is percentage of variance explained; *ESd* is Cohen's *d*; *P gain* is percentile gain of experimental group.

Of particular interest in Table 6.2 is the large effect size for home atmosphere (*ESd* = 1.42) and the comparatively low effect sizes for other more "popular" measures of SES such as income (*ESd* = .67), education (*ESd* = .38), occupation (*ESd* = .42), and their combined effects. About these findings, White notes:

> More striking, however, is the fact that measures of home atmosphere correlated much higher with academic achievement than did any single or combined group of the traditional indicators of SES. Recalling the comments by Jencks et al. (1972) cited earlier, there are many differences among families that can potentially affect the academic achievement of the children in addition to differences in education, occupational level, and income of the parents. It is not at all implausible that some low-SES parents (defined in terms of income, education, and/or occupational level) are very good at creating a home atmosphere that fosters learning (e.g., read to their children, help them with their homework, encourage them to go to college, and take them to the library and to cultural events), whereas other low-SES parents are not. (p. 471)

White concludes by noting that the real variable of interest in studies of influences on achievement might be best described as home environment. This provides for a much more optimistic perspective on SES than that considered from the perspective of previous research (e.g., Coleman and Jencks) or conventional wisdom. As the quotations above illustrate, the effects of SES are frequently thought of as impervious to change and extremely large. White's meta-analysis indicates that the effects are not as large as once thought. More important, if the ubiquitous SES effect is primarily a function of home environment, it can be altered. That is, interventions can be designed and implemented that provide parents with information and resources to establish a home environment that can positively affect students' academic achievement.

Prior Knowledge

Another apparent truism accepted by education practitioners and researchers is that prior knowledge is a strong determinant of academic achievement (see Alexander, Kulikowich, & Jetton, 1994; Bjorklund, 1985; Chi & Ceci, 1987; Chi, Glaser, & Farr, 1988; Glaser, Lesgold, & Lajoie, 1987; Pressley & McCormick, 1995; Schneider & Pressley, 1989). Table 6.3 lists effect sizes for prior knowledge as reported in various studies.

Table 6.3
Achievement and Prior Knowledge

Study	*ESd*	*P gain*	*PV*
Bloom (1976)	2.20[a]	48	54.76
Dochy (1992) (in Dochy, Segers, & Buehl, 1999)	1.71	46	42.25
Tobias (1994)	1.76[a]	46	43.56
Alexander, Kulikowich, & Schulze (1994)	1.04[a]	35	21.16
Dochy et al. (1999)	1.76[a]	46	43.56
Schiefele & Krapp (1996)	.43	16	4.41
Tamir (1996)	1.67	45	40.96
Boulanger (1981)	1.04	35	21.16
$\bar{x}$ ($Q = 69.4$, $df = 7$, $p < .05$)	1.43	42	33.64
$\bar{x}$ ($Q = 6.07$, $df = 4$, $p > .05$)	1.81	46	40.96

Note: Quantities were computed by beginning with the *r* reported in each study. These were transformed to *Zr* and an average was computed. The average *Zr* was then translated back to *r*. The *PV*, *ESd*, and *P gain* were then computed from the average *r*.
r is Pearson's product-moment correlation; *PV* is percentage of variance explained; *ESd* is Cohen's *d*; *P gain* is percentile gain of experimental group.
A *Q* statistic with $p < .05$ was interpreted as an indication that one or more correlations in the set were outliers. These outliers were identified using procedures described by Hedges and Olkin (1985). The *Q* statistic with outliers removed was then computed.
[a] Estimated from reported data.

Of the studies listed in Table 6.3, perhaps the most extensive was that conducted by Dochy, Segers, and Buehl (1999). In their analysis of 183 studies, Dochy et al. found that 91.5 percent of the studies demonstrated positive effects of prior knowledge on learning, and that those that did not measured prior knowledge in ways that were indirect, questionable, or even invalid. For example, some studies measured prior knowledge by simply asking students if they were familiar with a topic.

Table 6.3 reports findings from studies analyzing the effects of prior knowledge on academic achievement; however, prior knowledge also has been shown to be related to skills that might be considered "higher order" in nature. For example, Alexander and Judy (1994) focused their analysis on research related to the relationship between prior achievement and strategic or metacognitive knowledge. They identify the following generalizations about this relationship:

1. A foundation of domain-specific knowledge is necessary to acquire strategic knowledge.
2. Inaccurate or incomplete domain knowledge can exhibit the learning of strategic knowledge.
   Strategic knowledge contributes to the utilization and acquisition of domain-specific knowledge.
3. As knowledge in a domain increases, strategic knowledge is altered.
4. Differences in the relative importance of domain-specific and strategic knowledge may be a consequence of the nature of the domain or the structure of the task to which they are applied.

One study that is perhaps most illuminating relevant to this discussion is that conducted by Rolfhus and Ackerman (1999), even though it was not about prior knowledge, per se. Rolfhus and Ackerman helped define the structure of prior knowledge for academic subjects by assessing the domain-specific knowledge of 141 college students using traditional (i.e., forced-choice) tests for 20 academic domains. They then factor analyzed the correlations between those assessments. These results are reported in Table 6.4.

At least two elements of the findings reported in Table 6.4 are relevant to this discussion. First, and perhaps most striking, is the existence of a general factor that has factor loadings[1] greater than +.290 on all but one of the domain-specific tests (i.e., statistics). Yet, even this loading was .284. This implies that academic competence is grounded in a common core of knowledge, supporting arguments made by Hirsch (1996), Bennett (1992), and Finn (1991) that a strong general knowledge base enhances academic achievement.

A second relevant feature of the results reported in Table 6.4 is the existence of the four factors other than the general factor. As labeled in Table 6.4, they are (1) the humanities, (2) science, (3) civics, and (4) mechanics. If these factors represent commonalities between academic subjects, they might provide guidance in terms of organizing K–12 curricula. Specifically, the myriad of subjects currently addressed in most state curriculums via state-level content-area standards (see Marzano &

---

[1]A factor loading is an index of the relationship between a given measure — in this case, the various tests of achievement in the 20 academic subjects — and a latent construct represented by the factor. Generally, a factor loading of .300 or greater is interpreted as a significant relationship between a given measure and a given latent construct (see Mulaik, 1972). In this case, the .300 criterion was relaxed to .290 because a number of factor loadings were less than ten thousandths of a point within the .300 point criterion.

Kendall, 1999, for a discussion) might be organized into the four strands of humanities, science, civics, and mechanics as opposed to independent subject areas.

Table 6.4
The Factor Structure of Knowledge Tests

Test	Factor				
	General	**Humanities**	**Science**	**Civics**	**Mechanical**
Humanities					
American Literature	.612	.445			
Art	.367	.624			
Geography	.603				
Music	.551	.443			
World Literature	.665	.404			
Science					
Biology	.524	.359	.408		
Business/Management	.628		.330		
Chemistry	.426		.375		
Economics	.573	–.363	.387		
Physics	.556		.440		
Psychology	.526		.480		
Statistics					
Technology	.586		.318		
Civics					
American Government	.756			.299	
American History	.813			.344	
Law	.601				
Western Civilization	.705			.293	
Mechanics					
Astronomy	.508				.383
Electronics	.410				.425
Tools/Shop	.314				.625

Note: Adapted from "Assessing Individual Differences in Knowledge: Knowledge, Intelligence, and Related Traits," by E. L. Rolfhus and P. L. Ackerman, 1999, *Journal of Educational Psychology, 91*(3), p. 518.

Interest

Another student characteristic that presumably affects achievement is the interest students have in the content being learned. It makes great intuitive sense that if a student is not interested in a given topic, she will put little effort into the task of learning the content, and achievement will be affected. Table 6.5 presents findings from a number of studies that have examined the relationship between student interest and student achievement.

Table 6.5
Interest and Achievement

Study	*ESd*	*P gain*	*PV*
Schiefele, Krapp & Winteler (1992)	.63	24	9.00
Schiefele & Krapp (1995)	.75	27	12.25
Geisler-Brenstein & Schmeck (1996)	.93	32	17.65
Tobias (1994)	1.01	34	20.25
Bloom (1976)	.63	24	9.00
Steinkamp & Maehr (1983)	.39	15	3.61
$\bar{x}$ ($Q = 11.84$, $df = 5$, $p < .05$)	.73	27	11.56
$\bar{x}$ with outliers removed ($Q = 5.48$, $df = 4$, $p > .05$)	.80	29	13.69

Note: Quantities were computed by beginning with the r reported in each study. These were transformed to Zr and an average was computed. The average Zr was then transformed back to r.
r is Pearson's product-moment correlation; *PV* is percentage of variance explained; *ESd* is Cohen's d; *P gain* is percentile gain of experimental group.
A Q statistic with $p < .05$ was interpreted as an indication that one or more correlations in the set were outliers. These outliers were identified using procedures described by Hedges and Olkin (1985). The Q statistic with outliers removed was then computed.

As Table 6.5 shows, there is a moderate to strong relationship between interest and achievement for these studies; the average *ESd* is .80 when outliers are removed. Again, the dynamics of this relationship are fairly straightforward — the more interest students have in a topic, the more energy and attention they will put into the topic; consequently, the more they will learn about the topic. However, a number of studies have delved quite extensively into the working principles underlying this dynamic. For example, Schiefele and Csikszentmihalyi (1994) found that interest also correlates significantly with students' experience of efficacy, positive affect, and "potency" (feeling active, strong, and excited). An inference from their findings might be that the more students believe they can control a topic and have some say over how it is addressed and developed, the more interest they have in the topic. In their study, Alexander, Kulikowich, and Schulze (1994) found that as competence in a domain increases, there is a corresponding increase in one's interest in the domain. An inference here is that competence engenders interest, which in turn engenders more competence.

Aptitude

The final factor to consider within the general category of student background variables is aptitude. Again, there is a tacit assumption among educators and noneducators alike that aptitude or native ability plays a major role in achievement. Indeed, for decades arguments have been made that aptitude is the primary determiner of achievement. For example, Jensen (1980) and Heurnstein and Murray (1994) have argued that aptitude is not only the strongest predictor of academic achievement, but that it is a genetically determined, immutable characteristic. Table 6.6 lists the findings of a number of studies of the relationship between aptitude and achievement.

Table 6.6
Aptitude and Achievement

Study	*ESd*	*P gain*	*PV*
Fraser et al. (1987)	.88	31	16.00
Walberg (1984)	2.02	48	50.41
Bloom (1984a)	1.50	43	36.00
Dochy, Segers, & Buehl (1999)	.95	33	18.49
Bloom (1976)	1.62	45	39.69
Steinkamp & Maehr (1983)	.70	36	10.89
Boulanger (1981)	1.13	37	24.01
$\bar{x}$ ($Q = 52.02$, $df = 6$, $p < .05$)	1.25	39	28.09
$\bar{x}$ with outliers removed ($Q = 5.11$, $df = 2$, $p > .05$)	1.71	45	42.25

Note: Quantities were computed by beginning with the *r* reported in each study. These were transformed to *Zr* and an average was computed. The average *Zr* was then transformed back to *r*. The *PV*, *ESd*, and *P gain* were then computed from the average *r*.
r is Pearson's product-moment correlation; *PV* is percentage of variance explained; *ESd* is Cohen's *d*; *P gain* is percentile gain of experimental group.
A *Q* statistic with $p < .05$ was interpreted as an indication that one or more correlations in the set were outliers. These outliers were identified using procedures described by Hedges and Olkin (1985). The *Q* statistic with outliers removed was then computed.

The findings in Table 6.6 are fairly heterogeneous as indicated by the large value of the *Q* statistic when all estimates are considered as a group. To identify a set of homogeneous effect size estimates from which to compute an average estimate, four estimates were deleted. The average *ESd* with outliers excluded is 1.71.

One of the problematic aspects of much of the research on the relationship between aptitude and achievement is that measures of aptitude are frequently confounded with other student-level factors such as access to knowledge, interest, and so on. In fact, when the unique contribution to achievement attributable to aptitude is identified, it appears to be relatively small. To illustrate, consider the findings of Madaus et al. (1979), who found that the average correlation between achievement and aptitude (as measured by an IQ test) is .23 (ESd = .473) only when school-level, classroom-level, and home environment characteristics are partialed out and curriculum-specific dependent measures are used. The correlation between achievement and aptitude is .25 (ESd = .516) only when standardized tests are used.

Another problematic aspect of research in this area is defining exactly what is meant by aptitude. Although aptitude or intelligence can be described in a number of ways, one of the most widely accepted distinctions in the research and theory on intelligence is that between crystallized intelligence (*Gc*) and fluid intelligence (*Gf*). This distinction was first proposed by Cattell (1971/1987) and further developed by Ackerman (1996).

In brief, intelligence is thought of as consisting of two constructs: intelligence as knowledge (*Gc*, or crystallized intelligence) and intelligence as process (*Gf*, or fluid intelligence). Crystalized intelligence is exemplified by the ability to recognize or recall facts, generalizations, and principles along with the ability to learn and execute domain-specific skills and processes such as multiplying and dividing, reading, writing, and the like. Fluid intelligence is exemplified by procedures such as abstract reasoning ability, working memory capacity, and working memory efficiency. It is assumed that these mental processes are innate and not highly amenable to change through one's environment. Where fluid intelligence is assumed to be innate, crystalized intelligence is thought to be learned. However, it is also assumed that fluid intelligence is instrumental in the development of crystalized intelligence. That is, the more efficient one is at the cognitive processes involved in fluid intelligence, the more crystalized intelligence will be developed. A useful question relative to the present discussion is, What type of intelligence — crystalized or fluid — is more strongly related to academic achievement?

One of the most extensive studies of the relationship between *Gc*, *Gf*, and academic achievement was conducted by Rolfhus and Ackerman (1999). The researchers administered intelligence tests to 141 adults along with tests of knowledge in 20 different subject areas (discussed in the previous section of this chapter on prior knowledge). After factor-analyzing scores from nine subscales within the intelligence test, they found evidence for a general verbal factor, which they associated with *Gc,* and the existence of spatial and numeral factors, which they associated with *Gf*. To determine the relationship between *Gc* intelligence, *Gf* intelligence, and academic achievement, Rolfhus and Ackerman correlated the factor scores[2] with scores on the 20 academic domains. These results are reported in Table 6.7.

The most important point of Table 6.7 relative to the discussion is that in the domains of humanities, science, and civics, the verbal intelligence factor (*Gc*) has correlations greater than .200 with every achievement test except one (i.e., statistics) and correlations greater than .300 with over half of the achievement tests in these domains. Conversely, the two factors associated with *Gf* (the spatial and numerical factors) have no correlations greater than .300 with any of the achievement tests in any

[2]Factor scores are scores for individual subjects on the latent constructs (factors) identified within a factor analysis. When a set of tests is highly correlated, these factor scores are considered to be better estimates of the underlying traits that relate to the set of tests than are the individual scores on the tests themselves.

of the domains. (The spatial factor has no correlations greater than.200 with any of the achievement tests; thus, these correlations are not included in Table 6.7.) This implies that crystalized intelligence is a primary factor in the attainment of academic knowledge, where fluid intelligence is not. As stated by Rolfhus and Ackerman (1999), these findings suggest that academic "knowledge is more highly associated with *Gc*-type abilities than with *Gf*-type abilities" (p. 520). Taken as a whole, these findings appear to support the contention that "academic intelligence" is more a function of "learned knowledge" than of innate skills.

Table 6.7
Correlations Greater than .200 Between *Gc* and *Gf* Factor Scores and Tests of Academic Content

Test	Factor	
	Verbal (*Gc*)	**Numerical (*Gf*)**
Humanities		
American Literature	.432	
Art	.401	
Geography	.299	
Music	.404	
World Literature	.581	
Science		
Biology	.526	
Business/Management	.418	
Chemistry	.234	.282
Economics	.232	.204
Physics	.326	
Psychology	.381	
Statistics		
Technology	.305	
Civics		
American Government	.288	.255
American History	.317	
Law	.291	
Western Civilization	.394	
Mechanics		
Astronomy		.231
Electronics	.284	
Tools/Shop		

Note: Adapted from "Assessing Individual Differences in Knowledge: Knowledge, Intelligence, and Related Traits," by E. L. Rolfhus and P. L. Ackerman, 1999, *Journal of Educational Psychology, 91*(3), p. 520. Copyright © 1999 by the American Psychological Association. Adapted with the permission of APA and of Phillip L. Ackerman.

CONCLUSIONS ABOUT STUDENT-LEVEL VARIABLES

The research on student background variables presents a somewhat different and more optimistic picture than that usually ascribed to this literature base. Specifically, it appears that home environment is a more powerful predictor of student achievement than any other aspect of SES. Given that home environment is not a fixed trait, as is family or parental income, occupation, and the like, it might be the case that SES is more amenable to outside interventions than has been thought. Of all the student-level factors, prior knowledge has the largest effect on student achievement. This implies that the more students know about a topic, the more capable they are of learning new information about the topic. In addition, interest, which might be a function of competence, also influences achievement. Finally, the stronger relationship between crystalized intelligence and achievement than that between fluid intelligence and achievement also provides support for the hypothesis that "academic intelligence" is not a set of fixed traits impervious to change.

REVISITING THE THREE CATEGORIES

From the research reviewed in this and the previous two chapters, a case can be made that as a set, the three categories of variables have identifiable and somewhat stable influences on student achievement. Specifically, a case can be made that the percentage of variance accounted for by the three categories of variables are as follows:

student background:	80.00%
school level:	6.66%
teacher level:	13.34%

Again, it is important to note that these are conservative estimates from the perspective of the school- and classroom-level categories. That is, these estimates ascribe all variance that cannot as yet be attributed to school- or classroom-level variables to student background characteristics.

Based on the recommendations of Cohen and Cohen (1975) and Dawes and Currigan (1974), one might compute a viable estimate of the standardized regression coefficient[3] for these three predictors of achievement by using the *PV* for each set of variables. Using the standardized regression coefficients derived from the *PV*s reported above, predicted student achievement in *Z* score form would be expressed in the following way:

predicted achievement = .895 x student background + .365 x teacher characteristics + .257 x school characteristics

[3]Standardized regression coefficients are used when all scores — those for the predictor variables and those for the predicted variables — are expressed in *Z* score form. In this format, the regression coefficients are analogous to the partial correlations between the predictor variables and the predicted variables. In this case, it was assumed that the *PV*s for each predictor variable — student characteristics, teacher characteristics, school characteristics — represent the unique relationship between those variables and student achievement. Therefore, the regression weights (i.e., partial correlation coefficients) were estimated by computing the square root of each *PV*.

Using this equation, one can compute the predicted scores in Z score form[4] for various levels of student background, school characteristics, and classroom characteristics, as shown in Table 6.8.

Six situations are shown in Table 6.8:

- *Situation 1:* The achievement of students with an "average teacher" in an average school
- *Situation 2:* The achievement of students with an ineffective teacher in an ineffective school
- *Situation 3:* The achievement of students with an ineffective teacher in an exceptional school
- *Situation 4:* The achievement of students with an exceptional teacher in an ineffective school
- *Situation 5:* The achievement of students with an exceptional teacher in an exceptional school
- *Situation 6:* The achievement of students with an average teacher in an exceptional school

Conceptually, one might think of an exceptional teacher as one who makes optimum use of the teacher-level variables discussed in Chapter 5. More specifically, that teacher's use of these variables places him at the extreme positive end of the distribution of all teachers. The average teacher is one whose use of the teacher-level variables places him in the middle of the distribution, and the ineffective teacher is one whose use of the teacher-level variables places him at the extreme negative end of the distribution. The same interpretation can be applied to schools. The exceptional school is one whose use of the school-level variables places it at the extreme positive end of the distribution; an average school is in the middle of the distribution relative to its us of the school-level variables, and the ineffective school is at the extreme negative end of the distribution.

For each of the six situations included in Table 6.8, the predicted score for seven hypothetical students is presented. One student enters school with achievement in a particular subject that places him at –3.00 standard deviations — the student is performing at the extreme negative end of the distribution in that subject area. Another student enters the school performing at –2.00 standard deviations and another at –1.00 standard deviations. The student with an entrance Z score of 0 is performing precisely in the middle of the distribution. Finally, the next three students enter performing at +1.00, +2.00, and +3.00 standard deviations, respectively. In short, the seven hypothetical students broadly represent the range of student achievement in a given subject area.

---

[4] Z score form is standard score form. Observed scores are translated to Z score form using the following formula:

$$\boxed{\frac{x - \bar{x}}{SD}}$$

where x is the observed score, $\bar{x}$ is the mean of the set of scores, and SD is the standard deviation of the set of scores.

Table 6.8
Predicted Effects of School and Teacher on Student Achievement

	Student Achievement				Student Achievement		
Situation (School/Teacher)	**Enter[a]**	**Leave[b]**	**Net**	**Situation (School/Teacher)**	**Enter[a]**	**Leave[b]**	**Net**
#1	-3.0	-2.68	.32	*#4*	-3.0	-2.36	.64
Average School	-2.0	-1.79	.21	Ineffective School	-2.0	-1.47	.53
Average Teacher	-1.0	-.89	.11	Exceptional Teacher	-1.0	-.57	.43
	0	0	0		0	.32	.32
	+1.0	.89	-.11		+1.0	1.22	.22
	+2.0	1.79	-.21		+2.0	2.11	.11
	+3.0	2.68	-.32		+3.0	3.00	0
#2	-3.0	-4.55	-1.55	*#5*	-3.0	-.81	2.19
Ineffective School	-2.0	-3.66	-1.66	Exceptional School	-2.0	.08	2.08
Ineffective Teacher	-1.0	-2.76	-1.76	Exceptional Teacher	-1.0	.98	1.92
	0	-1.87	-1.87		0	1.87	1.87
	+1.0	-.98	-1.98		+1.0	2.76	1.76
	+2.0	-.08	-2.08		+2.0	3.66	1.66
	+3.0	.81	-2.19		+3.0	4.55	1.55
#3	-3.0	-3.00	0	*#6*	-3.0	-1.91	1.09
Exceptional School	-2.0	-2.11	-.11	Exceptional School	-2.0	-1.01	.99
Ineffective Teacher	-1.0	-1.22	-.22	Average Teacher	-1.0	-.12	.88
	0	-.32	-.32		0	.77	.77
	+1.0	.57	-.43		+1.0	1.67	.67
	+2.0	1.47	-.53		+2.0	2.56	.56
	+3.0	2.36	-.64		+3.0	3.46	.46

Note: The regression equation used to compute the values in Table 6.8 was predicted score = .895 x student background score + .365 x teacher score + .257 school score. Student, teacher, and school scores were conceptualized as a scale with a range of 0 to 10. An ineffective teacher was assigned a score of 0, an average teacher was assigned a score of 5, and an effective teacher was assigned a score of 10. Likewise, an ineffective school was assigned a score of 0, an average school was assigned a score of 5, and an effective school was assigned a score of 10. Thus, scores of 0 and 10 represent extremes. Additionally, these extreme scores were assigned *Z* scores of –3.00 (ineffective) and +3.00 (effective). The entire distribution of scores, then, was thought to span six standard deviations. Scores on the 0 to 10 scale were transformed in their *Z* score form and entered as values in the regression equation.
[a] The number of standard deviations (in *Z* score form) that a student's academic achievement is from the mean when he or she enters the school year.
[b] The number of standard deviations (in *Z* score form) that a student's academic achievement is from the mean when he or she leaves the school year.

The "Leave" column in each situation represents the predicted scores of the seven students in a specific subject area after a given period of time — for example, a school year.[5] The third column

[5]No precise statistic is available relative to the amount of time it takes for students to learn specific academic content. However, most of the studies that consider effects on student achievement look at achievement over a school year or less.

in each situation — "Net" — represents the net gain or loss in Z score units at the end of the school year. To illustrate, consider situation 1 (average school, average teacher) and the student who enters the class performing at the mean — his Z score is 0. That student will leave the year-long course performing at exactly the same place in terms of the distribution of student scores for that subject area — a Z score of 0. This is not to say that learning has not occurred. Indeed, recall Hattie's (1992) conclusions reported in Chapter 3. Specifically, Hattie estimated that one can expect an effect size (*ESd*) of .24 standard deviations due to maturation only. Our student entering the course performing at 0 standard deviations has learned, then, but he has not increased in standing relative to other students.

Table 6.8 paints an interesting picture of the influence of student background variables, teacher-level variables, and school-level variables on student achievement. Prior to discussing this picture, it is important to note that the predictions in Table 6.8 are based on the deductively inferred theoretical regression equation described earlier in this section. That model is surely a rough approximation only of the real-world relationships between student background variables, teacher-level variables, school-level variables, and academic achievement. The note at the end of this chapter describes some of the assumptions made in this model that may not mirror the real-world relationships among these categories of variables.

This caution aside, the figures listed in Table 6.8 are fodder for some thought-provoking hypotheses. They suggest that average schools and average teachers (situation 1), although they do little harm, do little to influence students' relative position on the distribution of achievement scores for all students. Those students who enter with relatively low standings exit with relatively low standings. Those who enter with relatively high standings exit with relatively high standings. Finally, the student who enters the course in the middle of the distribution ($Z = 0$) exits the course in the same position — the middle of the distribution.

The ineffective teacher in the ineffective school (situation 2) appears to have a negative impact on the standings of all students in his class. According to Table 6.8, the student who enters the class performing in the center of the score distribution ($Z = 0$), leaves the course with a Z score of –1.87. Even the student in the class of an ineffective teacher embedded in an exceptional school (situation 3) appears to lose ground in terms of relative achievement. The student who enters that teacher's course achieving in the middle of the score distribution ($Z = 0$) leaves performing below the mean of the distribution.

Situation 4 — the exceptional teacher in an ineffective school — produces some surprising results. All students either maintain their standing or increase it. The student entering the course performing at the middle of the distribution ($Z = 0$) leaves performing one-third of a standard deviation above the mean ($Z = .32$). Of course the exceptional teacher in the exceptional school (situation 5) produces the greatest gains in student achievement. The student entering the course in the center of the score distribution ($Z = 0$) exits performing almost two standard deviations above the mean ($Z = 1.87$). Finally, even the average teacher in an effective school (situation 6) produces positive effects. The student who enters the course performing in the center of the score distribution exits performing almost three-fourths of a standard deviation above the mean ($Z = .77$).

If valid, albeit tenuous, generalizations can be inferred from Table 6.8, one might be that "exceptional performance produces results." Exceptional performance in terms of school-level factors overcomes the average performance of teachers, but not the ineffective performance of teachers. However, exceptional performance on the part of teachers not only compensates for average performance at the school level, but even ineffective performance at the school level.

Chapter 6 Note:

At least two characteristics inherent in this theoretical model probably do not mirror real-world relationships. First, the model assumes that school-level, teacher-level, and student-level variables have independent relationships with academic achievement — there are no interaction effects represented in the model. Second, the predicted scores are subject to the statistical phenomenon of regression toward the mean as are all regression models. To illustrate the implications of this phenomenon, consider the predicted score for a student whose Z score on student-level characteristics is –3.00, and whose teacher-level and school-level Z scores are 0 representing an "average teacher" and an "average school." The regression equation for the student will be as follows:

$$
\begin{aligned}
\text{predicted } Z \text{ score} &= .895 \times (-3.00) + .360 \times (0) + .257 \times (0) \\
&= .895 \times (-3.00) \\
&= -2.68
\end{aligned}
$$

Thus, in the case of the predicted score for a student at the extreme end of the distribution in terms of achievement in an average school and with an average teacher, the predicted score is a function of the relationship between student background and achievement only. Since the relationship is not perfect, all predicted scores will be regressed toward the mean.

PART III:
APPLICATIONS

Chapter 7
USING THE KNOWLEDGE BASE ABOUT SCHOOL EFFECTIVENESS

Chapters 4, 5, and 6 attempted to establish the relative effects of three categories of variables influencing student achievement: school-level variables, classroom-level variables, and student-level variables. The key variables in each category of variables, as described in the previous three chapters, are reported in Table 7.1.

Table 7.1
Categories and Key Variables

Category	Key Variables
School	• Opportunity to learn • Time • Monitoring • Pressure to achieve • Parent involvement • School climate • Leadership • Cooperation
Teacher	• Instruction • Curriculum design • Classroom management
Student	• Home atmosphere • Prior knowledge • Aptitude • Interest

Given that Table 7.1 represents a fairly accurate accounting of the key variables within each of the three categories, a useful question is, How might educators use this information? This chapter considers three possible uses of this information: (1) as a model for staff development, (2) as a model for evaluation, and (3) as a model for data-driven school improvement.

STAFF DEVELOPMENT

As a model of staff development, the knowledge base about the three categories of variables would be used as the framework for a curriculum to be delivered to staff members in a school. To illustrate, Table 7.2 provides a brief description of the strategies that might be presented to staff members in a school for each of the variables in each of the three categories of variables.

Table 7.2
Strategies for Key Variables

Category	Variable	Strategies
School	Opportunity to learn	• strategies for aligning the curriculum and achievement tests • strategies for designing assessments aligned with the curriculum • strategies for ensuring that the curriculum is covered
	Time	• strategies for increasing the amount of allocated time • strategies for decreasing absenteeism and tardiness
	Monitoring	• strategies for setting school-wide achievement goals for students • strategies for collecting and reporting data on student achievement
	Pressure to achieve	• strategies for communicating the importance of students' academic achievement • strategies for celebrating and displaying student achievement
	Parental involvement	• strategies for involving parents in policy decisions • strategies for gaining parental support for policy decisions
	Climate	• strategies for identifying and communicating school rules and procedures • strategies for implementing and enforcing school rules and procedures
	Leadership	• strategies for articulating leadership roles • strategies for transferring and communicating key information • strategies for group decision making
	Cooperation	• strategies for developing consensus around key issues • strategies for increasing the frequency and quality of informal contacts among staff members • strategies for establishing and implementing behavioral norms among staff
Teacher	Instruction	• teaching strategies that ‣ enhance students' abilities to identify similarities and differences ‣ enhance students' abilities to summarize and take notes ‣ reinforce effort and provide recognition ‣ enhance the effectiveness of homework and practice ‣ enhance students' abilities to generate nonlinguistic representations ‣ provide students with opportunities to engage in cooperative learning ‣ enhance the effectiveness of academic goals and provide students with feedback ‣ enhance students' abilities to generate and test hypotheses ‣ activate students' prior knowledge

Category	Variable	Strategies
	Curriculum design	• planning strategies that ‣ enhance the manner in which instruction goals are ordered and paced within and between units ‣ enhance the manner in which instructional activities are ordered and paced within and between units
	Classroom management	• strategies that enhance the identification and implementation of rules and procedures for • room use • seatwork • group work • discipline
Student	Home atmosphere	• strategies for enhancing the extent to which parents provide their children with an environment that supports academic achievement
	Aptitude and prior knowledge	• strategies for enhancing students' general background knowledge
	Interest	• strategies for identifying and tapping into students' interests

Rather than present information about all of the strategies listed in Table 7.2, the most salient needs for a school would first be identified. This can be accomplished by collecting direct data on each element identified in Table 7.2 or by collecting perceptual data on each element. To illustrate, consider the school-level factor of time. A school could collect direct data on this factor by determining the actual amount of time allocated to instruction and the amount of time lost to absenteeism, or the school could collect data from teachers and administrators about their perceptions of the extent to which time was used effectively. The perceptual data are probably less accurate but easier to collect.

Regardless of the method used to collect data, those variables whose values are perceived as less than optimal would be targeted as the focus for staff development. This approach has been labeled the "rational decision-making model" (Sproull & Zubrow, 1981) in that it assumes that the three categories of variables have a straightforward, stable relationship with achievement in all schools. If a school can simply identify those variables on which it is not performing well, it can pinpoint and receive the information it needs to improve student achievement. As straightforward as this approach sounds, it has been severely criticized (see Murnane, 1987; Willms, 1992). To illustrate, Willms (1992) makes the following comments about this approach:

> Our knowledge about how schools have their efforts on instructional outcome is inadequate to support this kind of management strategy. . . . I doubt whether another two decades of research will . . .help us specify a *model for all seasons* — a model that would apply to all schools in all communities at all times. (p. 65)

EVALUATION

In the service of evaluation, the knowledge base about the three categories of variables developed in this monograph can be used to identify the achievement gain that can be associated with school- and teacher-level variables as opposed to student variables. In effect, an evaluation model seeks to "evaluate" schools by determining how much of the variance in student achievement in a particular school is attributable to school- and teacher-level variables as opposed to student background variables.

Perhaps the most well-known evaluation model is that used by Sanders and his colleagues, discussed briefly in Chapter 5. That evaluation model uses the linear equation described in Table 7.3.

Table 7.3
Wright et al.'s (1997) Linear Equation

Y = M + S + H + C + H*C + T (S*H*C*) + A*S + A*H + A*C + A*H*C + A*T (S*H) + E
Y is the gain score for an individual student M is the overall mean S is the school H is the level of heterogeneity (in achievement) at the classroom level C is the class size H*C is the heterogeneity-by-class size interaction T(S*H*C) is the teacher nested within a particular school (S) within a particular level of heterogeneity (H) within a particular class size (C) A is the achievement level (broken down into four groups) for the student A*S is the achievement-by-system interaction A*H is the achievement-by-heterogeneity interaction A*C is the achievement-by-class size interaction A*H*C* is the achievement-by-heterogeneity-by-class size interaction A*T(S*H*C) is the achievement-by-teacher interaction E is the error term

Note: See "Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation," by S. P. Wright, S. P., Horn, and W. L. Sanders, 1997, *Journal of Personnel Evaluation in Education, 11,* 57–67, specifically page 58.
All terms are considered fixed effects with the exception of T(S*H*C*), A*T(T*H*C), and E.

The model described in Table 7.3 allows for the determination of the effect of a particular teacher (*T*) within a particular school (*S*) with a specific level of in-class heterogeneity (*H*) for a specific class size (*C*). This effect is reflected in the term *T**(*S***H***C*) and can be examined while controlling for the effect of in-class heterogeneity (*H*), the achievement level of students (*A*), class size (*C*), the overall effects of the school (*S*), and the various interaction terms included in the model. The model could just as easily be used to evaluate the effect of the school (*S*) after partitioning out the effects of factors *A*, *C*, *H*, the nested teacher factor (*T*), and the various interaction terms explicit in the model.

A growing trend in the evaluation literature is the use of HLM (see Byrk & Raudenbush, 1992) to estimate the effects of various factors. HLM designs use "levels" of regression equations. To illustrate, consider the three-level hierarchical linear model described by Scheerens and Bosker, described in Table 7.4.

Table 7.4
Scheerens and Bosker's Three-Level Hierarchical Linear Model

$$Y_{ijk} = \beta_{0jk} + \beta_1 P_{ijk} + R_{ijk} \text{ (student level)}$$
$$\beta_{0jk} = \gamma_{00k} + \gamma_{001} T_{jk} + U_{0jk} \text{ (teacher level)}$$
$$\gamma_{00k} = \delta_{000} + \delta_{001} S_k + V_{00k} \text{ (school level)}$$

Y_{ijk} represents the achievement score of pupil *i* in a class taught by teacher *j* in school *k*.
β_{0jk} is the class-specific intercept.
γ_{00k} is the school-level intercept.
P is a student background variable.
T represents a teacher-level variable.
S_k represents a school-level variable.
R_{ijk} represents the error or residual term at the student level.
U_{0jk} represents the error or residual term at the teacher level.
V_{00k} represents the error or residual term at the school level.
β_1 is the regression coefficient representing the effect of the student background characteristics on achievement.
γ_{001} is the regression coefficient for the teacher variable.
δ_{001} is the regression coefficient for the school variable.
δ_{000} represents the grand mean.

Note: See *The Foundations of Educational Effectiveness* (p. 60), by J. Scheerens and R. J. Bosker, 1997, New York: Elsevier.

Using Scheerens and Bosker's model, there would be a unique student-level equation for every student in the study, a unique teacher-level equation for every teacher in the study, and a unique school-level equation for every teacher in the study. In the student-level equation, the achievement score of a specific pupil in a class taught by a specific teacher in a specific school (Y_{ijk}) is the sum of the class-specific intercept (β_{0jk}), the background characteristics of a specific student as measured on some scale (P_{ijk})[1] multiplied by the regression coefficient representing the effect of the student background characteristics on achievement (β_1), and all student-level variation not accounted for by the rest of the model (R_{ijk}). The teacher-level equation decomposes the class-specific intercepts (β_{0jk}) in the student-level equation. In effect, the teacher-level analysis seeks to account for the differences between class intercepts — differences in achievement from class to class. The school-level equation decomposes the school-level intercept (i.e., γ_{00k}) in the teacher-level equation.

[1]Commonly, the student background variable *P* is "centered" around the district mean. For example, if the background characteristic were SES as indicated by family income, a given student's score on this variable, *P*, would be centered by subtracting the average family income level in the district. This would render β_{0jk} the expected achievement score of a class where students exhibited average SES as measured by family income.

HLM analysis is generally preferred over the use of the nonhierarchical designs (i.e., Sanders' approach) when assessing the effects of teachers or schools in that a system of equations like that just described allows for the simultaneous estimation of

1. the effects of student background characteristics on the achievement of students nested within classes;
2. the effect of teacher-level variables on the achievement of individual classes nested within schools; and
3. the effect of school-level variables on the achievement of individual schools nested within a district.

Although these same effects might be estimated using a model that is not hierarchical, HLM provides for more precision in that error terms (i. e., R_{ijk}, U_{0jk}, V_{00k}) are computed for each level of analysis, whereas with non-HLM designs, the errors associated with students, teachers, and schools are confounded in a single term.

The knowledge base reviewed in this monograph might be used to improve the precision of evaluation models in that sets of student background variables, teacher-level variables, and school-level variables might be included in the evaluation equations and their impact on achievement accounted for. To illustrate, consider the student-level equation in the hierarchical model just discussed: $Y_{ijk} = \beta_{0jk} + \beta_1 P_{ijk} + R_{ijk}$. As described, the equation includes one student-level variable, P, with its associated regression weight, β_1. Given the discussion on student-level variables in Chapter 6, this equation could be expanded to include four student-level variables: home atmosphere (P_1), student prior knowledge of the content (P_2), student interest in the topic (P_3), and student aptitude (P_4). The student-level HLM equation would be $Y_{ijk} = \beta_{0jk} + \beta_1 P_{1ijk} + B_2 P_{2ijk} + B_3 P_{3ijk} + B_4 P_{4ijk} + R_{ijk}$.

The estimates of β_{0jk}, then, would represent the individual class means corrected for these four student-level variables. Comparing β_{0jk} terms within a school would be tantamount to comparing student achievement between classes for which the initial differences in four key student background variables had been accounted. The same type of comparison between schools could be made after teacher characteristics had been accounted for by including key teacher-level predictor variables in the teacher-level equation, and so on. In short, the knowledge base regarding student-, teacher-, and school-level variables allows for the specification of evaluation equations with more variables, which in turn leads to more precision in the evaluation of teachers, schools, and districts.

DATA-DRIVEN SCHOOL IMPROVEMENT

The final use of the knowledge base developed in this monograph is for "data-driven school improvement." Although it is clear that previous approaches to school improvement (see previous discussion of the staff development approach) do not take into consideration the unique features of specific teachers and specific schools, data-driven school improvement provides for just that. Using this approach, a school first determines the relationship between school-level variables, teacher-level variables, student-level variables, and student achievement. This might be done by applying an HLM model with multiple predictors at each level as described in the previous section. Data could be collected on student achievement and each predictor variable and their respective regression weights estimated. These regression weights would be considered"baseline" effects.

The schools and teachers involved in the data-driven school improvement effort would then identify specific school- and teacher-level innovations they believe have a high potential for enhancing student achievement. These innovations would be implemented for a specific period of time and then student achievement data would again be collected. To illustrate how data-driven school improvement might be used, consider the following scenario.

A district wishing to engage in a data-driven school reform effort first gathers information on each school regarding the following:

1. The extent to which the articulated curriculum is actually taught by teachers and covers key concepts in the state-level test
2. The extent to which instructional time is used effectively
3. The extent to which specific achievement goals are articulated and progress toward those goals monitored
4. The extent to which the school communicates a clear message that student achievement is a primary goal
5. The extent to which parents are involved in and support school policies
6. The extent to which an orderly atmosphere is established and maintained
7. The extent to which the school establishes and maintains a cooperative atmosphere
8. The extent to which leadership roles are clearly articulated and consensus is promoted

In addition, the district gathers data from each teacher on the following variables:

1. Use of effective instructional strategies
2. Use of effective management techniques
3. Effective unit planning

From students, data relative to the following variables are collected:

1. General aptitude
2. Family support for academics
3. Student interest in the topics presented at school
4. The general knowledge base of students

Of course, a district might elect not to gather data on all school-level, teacher-level, and student-level variables, opting instead to focus on those variables for which data are most easily obtained. For this discussion, however, we assume that a district wishes to collect data on all variables at each level.

As a baseline measure of "uncorrected" student learning, the district then has teachers administer teacher-designed pre- and posttests for a specific unit of instruction. The raw gain scores[2] from these tests are then regressed on the student-level variables using the student-level HLM equation, producing a mean gain score for each class (i.e., term β_{0jk}) that is conditional given the values of the student-level predictor variables. The teacher-level variables are used as predictors in the teacher-level equation, producing a mean gain score for each school (i.e., term γ_{00k}) that is conditional based on the values of the teacher-level predictor variables. The school-level variables are used as predictors in the school-level HLM equation, producing a district mean gain score (i.e., term δ_{000}) that is conditional based on the values of the school-level predictor variables. The three HLM terms, (β_{0jk}, γ_{00k}, and δ_{000}) could be considered estimates of the current status of teacher-, school-, and district-level achievement gain. These findings are used as the baseline information.

During the next phase of data-driven school improvement, individual teachers identify specific instructional, management, or planning strategies they would like to use in the hopes of improving their effectiveness at enhancing student learning. Similarly, individual schools identify specific strategies relative to their use of time, students' opportunity to learn, and so on. These teacher- and school-level strategies are then implemented for a specific period of time (e.g., a semester).

While the identified teacher- and school-level strategies are being implemented, pre- and posttest data are again collected for each student during a specific unit of instruction. At the end of the implementation period, data are again collected for teacher- and school-level predictor variables to reflect the implementation of teacher- and student-level strategies. The newly gathered achievement gain data are then regressed on student background variables, teacher-level variables, and school-level variables using the three levels of HLM equations. The terms within the new HLM equations (e.g., values for coefficients and values for variables) are compared to those in the baseline equations to determine the effectiveness of the teacher- and school-level interventions[3]. The cycle of data-driven school reform could be reinitiated on a yearly basis, thus allowing individual teachers, individual schools, and the district as a whole to engage in continuous data-driven school improvement from a "value-added" perspective.

---

[2]The use of gain scores is problematic in that they are characteristically less reliable than either the pretest or posttest scores from which they are derived (see Cohen & Cohen, 1975). However, this problem can be circumvented by entering pretest scores as predictor variables in the regression equation.

[3]One could logically conclude that an intervention used by a specific teacher increased student achievement if

a. the β_{0jk} term for that teacher increased, along with
b. an increase in the regression weight (e.g., γ_{001}) and an increase in the accompanying variable value (e.g., T_{jk}) for the intervention that was used, along with
c. a decrease in the error term (U_{0jk}) for that particular teacher.

Similarly, one could logically conclude that an intervention used by a specific school increased student achievement if

a. the γ_{00k} term for that school increased, along with
b. an increase in the regression weight (e.g., δ_{001}) and an increase in the accompanying variable value (e.g., S_k) for the intervention that was used, along with
c. a decrease in the error term (V_{00k}) for that particular school.

EPILOGUE

This monograph provides a quantitative review of the research literature regarding school-, teacher-, and student-level variables that affect student achievement. That review has resulted in a perspective on school reform that is far more optimistic than those promoted in the mid 1960s. In fact, the findings reviewed in this monograph indicate that schools can make profoundly influence student achievement. Specifically, the conclusions presented imply that student achievement can be strongly affected if schools

1. provide teachers with a well-articulated curriculum that specifically addresses the content on the assessments that are used to judge the academic achievement of students *and* ensure that the articulated curriculum is actually taught;
2. optimize their use of instructional time;
3. establish specific achievement goals for students and carefully monitor the extent to which those goals are being met;
4. communicate a clear message to all concerned that high academic achievement is the primary goal of the school;
5. involve parents in the processes of setting and enforcing policies;
6. maintain an orderly environment for all concerned;
7. maintain a cooperative environment for all concerned; and
8. involve staff in key decisions and establish clear lines of communication and leadership roles.

Along with the attention to the eight areas listed above, individual teachers must use the most effective instructional strategies, use the most effective managerial techniques, and design classroom curriculum effectively. This monograph also implies that not all school-level factors must be addressed equally or in the same way by each school. Similarly, not all teacher-level factors must be addressed equally by each teacher, implying that teacher improvement efforts will vary from teacher to teacher. In an effort to facilitate school- and teacher-specific improvement efforts, this monograph has articulated a process for data-driven improvement that uses some of the best available data analysis techniques to help make decisions about which interventions should be used by specific schools and specific teachers, and to evaluate the effectiveness of those interventions.

Finally, and perhaps most speculatively, the conclusions drawn in this monograph imply that even student background characteristics might be altered to some degree in a way that enhances student achievement. Specifically, student background might be altered by

- providing parents with information, resources, and techniques to make the home environment more conducive to academic achievement; and
- providing students with interventions aimed at increasing their understanding of a general academic knowledge base.

These suggestions are somewhat bold and highly optimistic given the perspective on school reform spawned in the 1960s. Even though there is a research base to support these recommendations, carrying them out will require the efforts of dedicated educators across all levels of the K–12 system.

REFERENCES

Abelson, R. P. (1985). A variance explained paradox: When a little is a lot. *Psychological Bulletin, 97,* 166–169.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22,* 227–257.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375–404.

Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts. *Review of Educational Research, 64*(2), 201–252.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *Review of Educational Research, 31*(2), 313–337.

Anania, J. (1982). The effects of quality of instruction on the cognitive and affective learning of students. (Doctoral dissertation, University of Chicago, 1981). *Dissertation Abstracts International, 42,* 4269A.

Anania, J. (1983). The influence of instructional conditions on student learning and achievement. *Evaluation in Education: An International Review Series, 7*(1), 1–92.

Austin, G. R. (1978). *Process evaluation: A comprehensive study of outliers*. Baltimore: Maryland State Department of Education.

Austin, G. R. (1979). *An analysis of outlier exemplary schools and their distinguishing characteristics*. College Park, MD: University of Maryland. Paper prepared for AERA annual meting, Los Angeles.

Austin, G. R. (1981). *Exemplary schools and their identification*. College Park, MD: University of Maryland, Center for Educational Research and Development. Unpublished paper.

Bennett, W. J. (1992). *The de-valuing of America: The fight for our culture and our children*. New York: Summit Books.

Berliner, D. C. (1979). Tempus educare. In P. L. Peterson & H. J. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications* (pp. 120–135). Berkeley, CA: McCutchan.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Bjorklund, D. F. (1985). The role of conceptual knowledge in the development of organization in children's memory. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development* (pp. 103–142). New York: Springer-Verlag.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Bloom, B. S. (1984a). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership, 41*(8), 4–18.

Bloom, B. S. (1984b). The 2 sigma problem: The search for methods of instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Boocock, S. S. (1972). *An introduction to the sociology of learning*. Boston: Houghton-Mifflin.

Borg, W. R. (1980). Time and school learning. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 33–72). Washington, DC: National Institute of Education, U.S. Department of Education.

Bosker, R. J. (1992). *The stability and consistency of school effects in primary education*. Enschede: University of Twente.

Bosker, R. J., & Witziers, B. (1995, January). *School effects, problems, solutions and a meta-analysis*. Paper presented at the International Congress for School Effectiveness and School Improvement, Leeuwarden, The Netherlands.

Bosker, R. J., & Witziers, B. (1996). *The magnitude of school effects. Or: Does it really matter which school a student attends?* Paper presented at the annual meeting of the American Educational Research Association, New York.

Boulanger, D. F. (1981). Ability and science learning. *Journal of Research in Science Teaching, 18*(2), 113–121.

Brookover, W. B., & Lezotte, L. W. (1979). *Changes in school characteristics coincident with changes in student achievement*. East Lansing: Institute for Research on Teaching, Michigan State University. (ERIC Document Reproduction Service No. ED 181 005)

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., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School social systems and student achievement: Schools can make a difference*. New York: Praeger.

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

Brophy, J. E., & Good, T. L. (1970). Teacher communication of differential expectations for children's classroom performance. *Journal of Educational Psychology, 61,* 365–374.

Bruggencate, C. G., Pelgrum, W. J., & Plomp, T. (1986). First results of the Second IEA Science Study in the Netherlands. In W. J. Nijhof & E. Warries (Eds.), *Outcomes of education and training*. Lisse: Swets & Zeitlinger.

Burke, A. J. (1984). Students' potential for learning contrasted under tutorial and group approaches to instruction (Doctoral dissertation, University of Chicago, 1983). *Dissertation Abstracts International, 44,* 2025A.

Byrk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. New York: Sage.

Cahen, S., & Davis, D. (1987). A between-grade levels approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal, 24,* 1–2.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 723–733.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action* (Rev. ed.). Amsterdam: North Holland Press. (Original work published 1971)

Charters, W. W., Jr. (1963). The social background of teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (Ch. 14). Chicago: Rand McNally.

Chi, M. T. H., & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 91–142). Orlando, FL: Academic Press.

Chi, M. T. H., Glaser, R., & Farr, M. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Clark, C. M., & Yinger, R. J. (1979). Teacher's thinking. In P. L. Peterson & H. J. Walberg (Eds.), *Research in teaching: Concepts, findings and implications* (pp. 231–263). Berkeley, CA: McCutchan.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *American Economic Review* (supplement), 139–165.

Cohen, J. (1988). *Statistical power for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: John Wiley & Sons.

Cohen, M. (1981). Effective schools: What the research says. *Today's Education, April/May,* 466–496.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U. S. Government Printing Office.

Cooper, H. M. (1989). *Homework*. White Plains, NY: Longman.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87,* 442–449.

Cotton, K. (1995). *Effective schooling practices: A research synthesis. 1995 update*. School Improvement Research Series. Portland, OR: Northwest Regional Educational Laboratory.

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

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95–100.

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

Dochy, F. (1992). *Assessment of prior knowledge as a determinant for future learning*. London: Jessica Kingsley. (http://ericae.net/books/dochy1/)

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relationship between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*(2), 145–186.

Edmonds, R. R. (1979a). *A discussion of the literature and issues related to effective schooling*. Cambridge, MA: Center for Urban Studies, Harvard Graduate School of Education.

Edmonds, R. R. (1979b, October). Effective schools for the urban poor. *Educational Leadership, 37*, 15–27.

Edmonds, R. R. (1979c). Some schools work and more can. *Social Policy, 9,* 28–32.

Edmonds, R. R. (1981a). Making public schools effective. *Social Policy, 12*(2).

Edmonds, R. R. (1981b). *A report on the research project, "Search for effective schools..." and certain of the designs for school improvement that are associated with the project*. Unpublished report prepared for NIE. East Lansing, MI: The Institute for Research on Teaching, College of Education, Michigan State University.

Edmonds, R. R., & Frederiksen, J. R. (1979). *Search for effective schools: The identification and analysis of city schools that are instructionally effective for poor children*. (ERIC Document Reproduction Service No. ED 170 396)

Elberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economic Education Review, 1*, 291–299.

Emmer, E. T., Evertson, C. M., Sanford, J. P., Clements, B. S., & Worsham, M. E. (1984). *Classroom management for secondary teachers*. Englewood Cliffs, NJ: Prentice Hall.

Evertson, C. M., Emmer, E. T., Clements, B. S., Sanford, J. P., & Worsham, M. E. (1984). *Classroom management for elementary teachers*. Englewood Cliffs, NJ: Prentice Hall.

Finn, C. E., Jr. (1991). *We must take charge: Our schools and our future*. New York: The Free Press.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Synthesis of educational productivity research [Special issue]. *Journal of Educational Research, 11*(2).

Geisler-Brenstein, E., & Schmeck, R. R. (1996). The revised inventory of learning processes: A multifaceted perspective on individual differences in learning. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge* (pp. 284–317). Boston: Kluwer.

Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing and measurement* (pp. 41–85). Hillsdale, NJ: Erlbaum.

Glass, G. V. (1976). Primary, secondary, and meta-analyses of research. *Educational Researcher, 5*, 3-8.

Glenn, B. C. (1981). *What works? An examination of effective schools for poor black children*. Cambridge, MA: Center for Law and Education, Harvard University.

Goldstein, H. (1997). Methods of school effectiveness research. *School Effectiveness and School Improvement, 8*(4), 369–395.

Good, T. L., & Brophy, J. E. (1986). School effects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 570–602). New York: Macmillan.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education, 36*(1), 5–13.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*(2), 99–136.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Heurnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: The Free Press.

Hirsch, E. D., Jr. (1996). *The schools we need and why we don't have them*. New York: Doubleday.

Horn, A., & Walberg, H. J. (1984). Achievement and interest as a function of quantity and quality of instruction. *Journal of Educational Research, 77,* 227–237.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter, M. (1984). Knowing, teaching, and supervising. In P. L. Hosford (Ed.), *Using what we know about teaching* (pp. 169–192). Alexandria, VA: Association for Supervision and Curriculum Development.

Husen, T. (1967). *International study of achievement in mathematics: A comparison of twelve countries*. New York: Wiley.

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

Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.

Johnson, D., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89*(1), 47–62.

Johnson, E. G., & Zwick, R. (1990). *Focusing the new design: The NAEP 1988 technical report: The nation's report card*. Princeton, NJ: Educational Testing Service.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Erlbaum.

Kifer, E. (2000). *Large-scale assessment: Dimensions, dilemmas, and policy*. Thousand Oaks, CA: Corwin Press.

Klitgaard, R. E., & Hall, G. R. (1974). Are there unusually effective schools? *Journal of Human Resources, 74,* 90–106.

Kulik, J. A., & Kulik, C. L. C. (1989). Meta-analysis in education. *International Journal of Educational Research, 13*(3), 221–340.

Lezotte, L. W., Edmonds, R., & Ratner, G. (1974, September). *A final report: Remedy for school failure to equitably deliver basic school skills*. East Lansing, MI: Department of Urban and Metropolitan Studies, Michigan State University.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist, 48*(12), 1181–1209.

Luyten, H. (1994). *School effects: Stability and malleability*. Enschide: University of Twente.

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

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

Manassee, A. L. (1985). Improving conditions for principal effectiveness: Policy implications of research. *Elementary School Journal, 85,* 439–463.

Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, CO: Mid-continent Research for Education and Learning (Document Reproduction No. ED 427 087)

Marzano, R. J., Gaddy, B. B., & Dean, C. (2000). *What works in classroom instruction*. Aurora, CO: Mid-continent Research for Education and Learning.

Marzano, R. J., & Kendall, J. S. (with B. B. Gaddy). (1999). *Essential knowledge: The debate over what American students should know*. Aurora, CO: Mid-continent Regional Educational Laboratory.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

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

Mosteller, F., & Moynihan, D. P. (1972). A pathbreaking report. In F. Mosteller & D. P. Moynihan (Eds.), *On equality of educational opportunity*. New York: Vintage Books.

Mulaik, S. A. (1972). *The foundation of factor analysis*. New York: McGraw-Hill.

Murnane, R. J. (1987). Improving education indicators and economic indicators: The same problems? *Education Evaluation and Policy Analysis, 9*(2), 101–116.

New York State Department of Education. (1974a, March). *Reading achievement related to educational and environmental conditions in 12 New York City elementary schools*. Albany, NY: Division of Education Evaluation.

New York State Department of Education. (1974b). *School factors influencing reading achievement: A case study of two inner city schools*. Albany, NY: Office of Education Performance Review. (ERIC Document Reproduction Service No. 089 211)

New York State Department of Education. (1976, March). *Three strategies for studying the effects of school processes*. Albany, NY: Bureau of School Programs Evaluation.

Nuthall, G. (1999). The way students learn. Acquiring knowledge from an integrated science and social studies unit. *Elementary School Journal, 99*(4), 303–341.

Nuthall, G., & Alton-Lee, A. (1995). Assessing classroom learning. How students use their knowledge and experience to answer classroom achievement test questions in science and social studies. *American Educational Research Journal, 32*(1), 185–223.

Odden, A. (1982, summer). Schools can improve: Local strategies need state backing. *State Education Leader,* 1–3.

Pelgrum, W. J., Eggen, Th. J. H. M.,& Plomp, Tj. (1983) *The second mathematics study: Results*. Enschede: Twente University.

Pressley, M., & McCormick, C. B. (1995). *Advanced educational psychology for educators, researchers and policymakers*. New York: Harper Collins College Publishers.

Purkey, S. C., & Smith, M. S. (1982). *Effective schools: A review*. Madison, WI: Wisconsin Center for Educational Research, School of Education, University of Wisconsin at Madison.

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

Ralph, J. H., & Hennessey, J. (1983, June). Science or reform: Some questions about the effective schools model. *Phi Delta Kappan,* 689–694.

Raudenbush, S. W., & Byrk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of Research in Education: Vol. 15* (pp. 423–475). Washington, DC: American Educational Research Association.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*(4), 307–335.

Reynolds, M. C., Wang, M. C., & Walberg, H. J. (1992). The knowledge base for special and general education. *Remedial and Special Education, 13*(5), 6–10.

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology, 91*(3), 511–526.

Rosenshine, B. V. (1979). Content, time, and direct instruction. In P. L. Peterson & H. J. Walberg (Eds.), *Research on teaching* (pp. 28–56). Berkeley, CA: McCutchon.

Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*(2), 166–169.

Rovee-Collier, C. (1995). Time windows in cognitive development. *Developmental Psychology, 31*(2), 147–169.

Rowe, K. J., & Hill, P. W. (1994). *Multilevel modeling in school effectiveness research: How many levels*. Paper presented at the International Congress for School Effectiveness and Improvement, Melbourne, Australia.

Rowe, K. J., Hill, P. W., & Holmes-Smith, P. (1995). Methodological issues in educational performance and school effectiveness research: A discussion with worked examples. *Australian Journal of Education, 39*(3), 217–248.

Rutter, M., Maughan, B., Mortimer, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. London: Open Books.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299–311.

Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London: Cassell.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. New York: Elsevier.

Schiefele, U., & Csikszentmihalyi, M. (1994). Interest and the quality of experience in classrooms. *European Journal of Psychology of Education, 9,* 251–270.

Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences, 8*(2), 141–160.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183–212). Hillsdale, NJ: Erlbaum.

Schneider, W., & Pressley, M. (1989). *Memory development between 2 and 20*. New York: Springer-Verlag.

Spartz, J. L. (1977). *Delaware educational accountability system case studies: Elementary schools grades 1–4*. Dover, DE: Delaware Department of Public Instruction.

Sproull, L. S., & Zubrow, D. (1981). Performance information in school systems: Perspectives from organizational theory. *Educational Administration Quarterly, 17*(3), 61–79.

Steinkamp, M. W., & Maehr, M. L. (1983). Affect, ability, and science achievement: A quantitative synthesis of correlational research. *Review of Educational Research, 53*(3), 369–396.

Stringfield, S., & Teddlie, C. (1989). The first three phases of the Louisiana school effectiveness study. In B. P. M. Creemers, T. Peters, & D. Reynolds (Eds.), *School effectiveness and school improvement: Proceedings of the Second International Congress, Rotterdam* (pp. 281–294). Lisse: Swets & Zeitlinger.

Tamir, P. (1996). Science assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge* (pp. 93–129). Boston: Kluwer.

Teddlie, C., Reynolds, D., & Sammons, P. (2000). The methodology and scientific properties of school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 55–133). London: Falmer Press.

Tobias, S. (1994). Interest, prior knowledge and learning. *Review of Educational Research, 64*(1), 37–54.

van der Werf, G. (1997). Differences in school and instruction characteristics between high-, average-, and low-effective schools. *School Effectiveness and School Improvement, 8*(4), 430–448.

Venezky, R. L., & Winfield, L. F. (1979). *Schools that succeed beyond expectations in reading*. (Studies on Educational Technical Report No. 1). Newark: University of Delaware. (ERIC Document Reproduction Service No. ED 177 484)

Walberg, H. J. (1980). *A psychological theory of educational productivity*. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED 206 042)

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership,* 19–27.

Walberg, H. J. (1999). Productive teaching. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (pp. 75–106). Berkeley, CA: McCutchan Publishing Corporation.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249–294.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1995). Learning influences. In H. J. Walberg & G. D. Haertel (Eds.), *Psychology and educational practice* (pp. 199–211). Berkeley, CA: McCutchan.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research, 84,* 30–43.

Wang, M. C., Reynolds, M. C., & Walberg, H. J. (1994). What works and what doesn't work: The case for the inclusive system. In K. K. Wong & M. C. Wang (Eds.), *Rethinking policy for at-risk students* (pp. 151–176). Berkeley, CA: McCutchan.

Wang, M. C., Reynolds, M. C., & Walberg, H. J. (Eds.). (1987–1991). *Handbook of special education: Research and practice* (Vols. 1–4). Oxford, England: Pergamon.

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

Welch, F. (1974). Relationships between income and schooling. In F. N. Kerlinger (Ed.), *Review of research in education*. Ithaca, NY: Peacock.

White, K. R. (1982). The relationship between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461–481.

Williams, B. I., Richmond, P. A., & Mason, B. J. (Eds.). (1986). *Designs for compensatory education: Conference proceedings and papers*. Washington, DC: Research and Evaluation Associates.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: The Falmer Press.

Wittrock, M. C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Witziers, B., & Bosker, R. J. (1997, January). *A meta-analysis on the effects of presumed school effectiveness enhancing factors*. Paper presented at ICSEI, Memphis, TN.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement. Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57–67.